

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;
 DR WPI: 2001-639232/73.
 XX
 XX
 PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 PS Example 17; Page 395; 579pp; English.

CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumor protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is a prostate specific
 CC peptide of the invention. The peptides either represent antigenic
 CC epitopes or domains of prostate specific proteins.

SO Sequence 30 AA:

Query Match 100.0%; Score 148; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2,7e-15;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGFAIPIGAMAIAGOI 30
 DB 1 taasdnfqlsgggfaipligamalaagqi 30

RESULT 2
 AAM01143
 ID AAM01143 standard; Protein; 30 AA.
 XX
 AC AAM01143;
 XX
 DT 04-OCT-2001 (first entry)

DE Mycobacterium tuberculosis antigen Ra12 first 30 amino acids.
 XX
 DE Human; prostate cancer; prostate-specific; diagnosis; vaccine;
 KM cytostatic; gene therapy; metastasis.
 XX

OS Mycobacterium tuberculosis.
 XX
 PN WO200151633-A2.
 XX

PD 19-JUL-2001.
 XX
 PE 16-JAN-2001; 2001WO-US01574.
 XX
 PR 14-JAN-2000; 2000US-0483672.
 XX

PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
 PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skelky YAW;
 PI Wang A, Heagner MJ;
 XX

DR WPI: 2001-425873/45.
 XX
 XX New polynucleotide encoding a prostate-specific protein, for
 PT diagnosing, monitoring and treating prostate cancer in a patient and
 PT for use in vaccines -
 XX

PS Example 17; Page 394; 543pp; English.

CC The present invention describes polynucleotide sequences (I) which encode
 CC prostate-specific proteins (II), (I) and (II) have cytostatic activity,
 CC and can be used in vaccine production and gene therapy. (I), (II),
 CC antibodies to (II), fusion proteins comprising (II), and isolated
 CC T cells prepared using (I) or (II) are used treat cancer in a patient.
 CC (I) and the antibodies are also used in the detection of cancer in a
 CC patient. The cancer that is diagnosed or treated is particularly
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
 CC (I) can be used for monitoring the progression of cancer in a patient.
 CC (I) and (II) can also be used to improve diagnostic and therapeutic
 CC methods for prostate cancer. They can indicate the level of metastasis
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
 CC AAM01318 represent polynucleotide and amino acid sequences used in the
 CC exemplification of the present invention.

SO Sequence 30 AA:

Query Match 100.0%; Score 148; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2,7e-15;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGFAIPIGAMAIAGOI 30
 DB 1 taasdnfqlsgggfaipligamalaagqi 30

RESULT 3
 AAG99028
 ID AAG99028 standard; Protein; 30 AA.
 XX
 AC AAG99028;
 XX
 DT 25-SEP-2001 (first entry)

DE Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.
 XX
 DE Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
 KM chromosome 22q11.2; prostate-specific protein; chromosome 1;
 KM prostate specific antigen; PSA.
 XX

OS Mycobacterium tuberculosis.
 XX
 PN WO200134802-A2.
 XX
 PD 17-MAY-2001.
 XX

PE 09-NOV-2000; 2000WO-US30904.
 XX
 PR 12-NOV-1999; 99US-0439313.
 XX
 PR 18-NOV-1999; 99US-0443686.
 XX

PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
 PI Kalos MD, Retter MW, Stolk JA, Day CH, Skelky YAW, Wang A;
 XX
 DR WPI: 2001-308785/32.
 XX

PT Isolated polypeptide comprising at least an immunogenic portion of a
 PT prostate-specific protein, useful in the diagnosis and therapy of
 PT prostate cancer -
 PS Example 17; Page 294; 325pp; English.

CC The present invention describes an isolated polypeptide (PI) comprising
 CC at least an immunogenic portion of a prostate-specific protein, or its
 CC variant. Also described are polynucleotides (NI) encoding (PI). (PI) and
 CC (NI) have cytostatic activity and can be used in vaccine production.
 CC The polypeptides, nucleic acids and antibodies from the present
 CC invention are useful in the diagnosis and therapy of prostate cancer.


```

AAM81657
ID AAM81657 standard; Protein; 132 AA.
XX
AC AAM81657;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TbrA12.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PE 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR N-PSDB; AAV64450.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Example 3; Page 97-98; 230pp; English.
XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC for inducing protective immunity against tuberculosis (TB). This sequence
CC can be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 132 AA;

```

Query Match 100.0%; Score 148; DB 19; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TAASDNFQLSGGGQGFPAIPGQAMAIAGQI 30
   |||||
Db 1 taasdnfqlsggqgfaipligamataigqi 30

```

RESULT 7
ID AAM64294 standard; Protein; 132 AA.
XX
AC AAM64294;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbrA12.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbrA12.
XX
OS Mycobacterium tuberculosis strain H37Ra.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.

```

XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR N-PSDB; AAV44342.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 101; 250pp; English.
XX
CC This polypeptide comprises Mycobacterium tuberculosis soluble
CC antigen TbrA12. It is encoded by a DNA sequence (see AAV44342)
CC isolated from a M. tuberculosis strain H37Ra expression library
CC with rabbit anti-sera raised against M. tuberculosis supernatant.
CC The invention relates to compositions and methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAM64291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
CC as well as DNA sequences encoding such polypeptides, recombinant
CC expression vectors and transformed or transfected host cells. Also
CC claimed are methods and diagnostic kits for detecting M.
CC tuberculosis infection in a patient using the above polypeptides,
CC antibodies, or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 132 AA;

```

Query Match 100.0%; Score 148; DB 19; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TAASDNFQLSGGGQGFPAIPGQAMAIAGQI 30
   |||||
Db 1 taasdnfqlsggqgfaipligamataigqi 30

```

RESULT 8
ID AAY39096 standard; Protein; 132 AA.
XX
AC AAY39096;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen TbrA12 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PE 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.

```

XX  Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI  Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX  WPI: 1999-527409/44.
DR  N-PSDB; AA219252.
XX
PT  New antigens from Mycobacterium tuberculosis useful in diagnostic
PT  skin tests and protective or therapeutic vaccines or compositions
XX
PS  Example 3; Page 98; 299pp; English.
XX
CC  The present invention describes polypeptides comprising an immunogenic
CC  part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC  are vaccines and fusion protein containing M. tuberculosis Ag's,
CC  M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC  other polypeptides fragments, can be used in pharmaceutical compositions
CC  or vaccines to generate a protective or therapeutic immune response to
CC  M. tuberculosis and as reagents in skin tests for diagnosis of
CC  tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC  by, T, B or natural killer cells and/or macrophages in
CC  tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
CC  AA219225 are used in the exemplification of the present invention.
XX
SQ  Sequence 132 AA;

Query Match          100.0%; Score 148; DB 20; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 TAAADNFOLSGGGGFAIPIGOMAIAGOI 30
    |||
    1 taasdntqlsggggfaipigamalaqgi 30

RESULT  9
AA219252
ID  AA219252 standard; Protein; 132 AA.
XX
AC  AA219252;
XX
DT  05-NOV-1999 (first entry)
XX
DE  M. tuberculosis recombinant antigen protein Tbra12.
XX
KW  Antigen; diagnosis; detection; infection; antibody; immunisation;
XX  vaccine; immunity.
XX
OS  Mycobacterium tuberculosis.
XX
PN  WO9942118-A2.
XX
PD  26-AUG-1999.
XX
PE  17-FEB-1999; 99WO-US03265.
XX
PR  05-MAY-1998; 98US-0072596.
XX  18-FEB-1998; 98US-0024753.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI  Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
XX  WPI: 1999-527416/44.
XX  N-PSDB; AA219040.
XX
XX  New polypeptide comprising antigenic portions of M. tuberculosis
XX  Example 3; Page 138; 323pp; English.
XX
XX  This invention describes novel recombinant antigens and their encoding

```

```

CC  nucleic acids derived from Mycobacterium tuberculosis. The novel
CC  polypeptides are useful for detecting M. tuberculosis infection in a
CC  biological sample by detecting antibodies which bind with the
CC  polypeptides, and are useful as vaccines for immunizing against
CC  M. tuberculosis infection. The new detection methods are needed as
CC  current vaccination strategies do not provide 100% immunity.
XX
SQ  Sequence 132 AA;

Query Match          100.0%; Score 148; DB 20; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 TAAADNFOLSGGGGFAIPIGOMAIAGOI 30
    |||
    1 taasdntqlsggggfaipigamalaqgi 30

Db  1 taasdntqlsggggfaipigamalaqgi 30

RESULT  10
AAU69898
ID  AAU69898 standard; Protein; 132 AA.
XX
AC  AAU69898;
XX
DT  30-JAN-2002 (first entry)
XX
DE  Mycobacterium tuberculosis antigen Ra12.
XX
KW  Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
XX  fusion protein; Ra12 antigen.
XX
OS  Mycobacterium tuberculosis.
XX
PN  WO200173032-A2.
XX
PD  04-OCT-2001.
XX
PE  27-MAR-2001; 2001WO-US09919.
XX
PR  27-MAR-2000; 2000US-0536857.
XX  09-MAY-2000; 2000US-0568100.
XX  12-MAY-2000; 2000US-0570737.
XX  13-JUN-2000; 2000US-0593793.
XX  27-JUN-2000; 2000US-0605783.
XX  10-AUG-2000; 2000US-0636215.
XX  29-AUG-2000; 2000US-0651236.
XX  06-SEP-2000; 2000US-0657279.
XX  02-OCT-2000; 2000US-0679426.
XX  10-OCT-2000; 2000US-0685166.
XX
XX  (CORI-) CORIXA CORP.
XX
XX  Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
XX  Fanger GR, Reller MW, Stolk JA, Day CH, Vedvick TS, Carter D;
XX  Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
XX
XX  WPI: 2001-639232/73.
XX
XX  New human prostate-specific polypeptides and polynucleotides useful for
XX  the diagnosis and treatment of cancer, especially prostate cancer -
XX
XX  Example 17; Page 531-532; 579pp; English.
XX
XX  The invention relates to isolated prostate-specific
XX  polynucleotides, polypeptides, fusion proteins of the polypeptides,
XX  antibodies raised against the polypeptides (or antigenic epitopes
XX  derived from them) and antigen-presenting cells expressing the
XX  polypeptides. The antibodies are useful for detecting the presence of
XX  cancer, especially prostate cancer. The polypeptides, polynucleotides
XX  and the antigen-presenting cells are useful for stimulating and/or expanding
XX  T cells specific for a tumour protein, and for inhibiting the development
XX  of cancer especially prostate cancer. Compositions comprising the

```

CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.

XX Sequence 132 AA;

Query Match 100.0%; Score 148; DB 22; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1,4e-14;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFQLSGGGGFAIPIGAMAIAGOI 30
 |||||
 Db 1 taasdnfqlsggggfaipigamalaiagqi 30

RESULT 11

ID AA069906 standard; Protein; 132 AA.

XX AA069906;

XX 30-JAN-2002 (first entry)

XX Mycobacterium tuberculosis antigen Ra12.

XX Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;

XX fusion protein; Ra12 antigen.

XX Mycobacterium tuberculosis.

XX WO200173032-A2.

XX 04-OCT-2001.

XX 27-MAR-2001; 2001WO-US09919.

XX 27-MAR-2000; 2000US-0536857.

XX 09-MAY-2000; 2000US-0568100.

XX 12-MAY-2000; 2000US-0570737.

XX 13-JUN-2000; 2000US-0593793.

XX 27-JUN-2000; 2000US-0605783.

XX 10-AUG-2000; 2000US-0636215.

XX 29-AUG-2000; 2000US-0651236.

XX 06-SEP-2000; 2000US-0657279.

XX 02-OCT-2000; 2000US-0679426.

XX 10-OCT-2000; 2000US-0685166.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stoik JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
 XX WPI; 2001-639232/73.

XX New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 XX Example 17; Page 541-542; 579pp; English.

XX The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune

CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.

XX Sequence 132 AA;

Query Match 100.0%; Score 148; DB 22; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1,4e-14;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASDNFQLSGGGGFAIPIGAMAIAGOI 30
 |||||
 Db 1 taasdnfqlsggggfaipigamalaiagqi 30

RESULT 12

ID AAM01253 standard; Protein; 132 AA.

XX AAM01253;

XX 04-OCT-2001 (first entry)

XX Mycobacterium tuberculosis antigen Ra12 amino acids.

XX Human, prostate cancer; prostate-specific; diagnosis; vaccine;

XX cytostatic; gene therapy; metastasis.

XX Mycobacterium tuberculosis.

XX WO200151633-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001WO-US01574.

XX 14-JAN-2000; 2000US-0483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
 PI Kalos MD, Fanger GR, Day CH, Retter MW, Stoik JA, Skeiky YAW,
 PI Wang A, Meagher MJ;
 XX WPI; 2001-425873/45.

XX New polynucleotide encoding a prostate-specific protein, for
 PT diagnosing, monitoring and treating prostate cancer in a patient and
 PT for use in vaccines -
 XX Example 17; Page 492; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode
 CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
 CC and can be used in vaccine production and gene therapy. (II), (III),
 CC antibodies to (II), fusion proteins comprising (II), and isolated
 CC T cells prepared using (I) or (II) are used treat cancer in a patient.
 CC (I) and the antibodies are also used in the detection of cancer in a
 CC patient. The cancer that is diagnosed or treated is particularly
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
 CC (I) can be used for monitoring the progression of cancer in a patient.
 CC (I) and (II) can also be used to improve diagnostic and therapeutic
 CC methods for prostate cancer. They can indicate the level of metastasis
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
 CC AAM01318 represent polynucleotide and amino acid sequences used in the
 CC exemplification of the present invention.

XX Sequence 132 AA;

XX Query Match 100.0%; Score 148; DB 22; Length 132;


```

XX      WO200173032-A2.
PN
XX      04-OCT-2001.
PD
XX      27-MAR-2001; 2001WO-US09919.
PF
XX      27-MAR-2000; 2000US-0536857.
PR      09-MAY-2000; 2000US-0568100.
PR      12-MAY-2000; 2000US-0570737.
PR      13-JUN-2000; 2000US-0593793.
PR      27-JUN-2000; 2000US-0605783.
PR      10-AUG-2000; 2000US-0636215.
PR      29-AUG-2000; 2000US-0651236.
PR      06-SEP-2000; 2000US-0657279.
PR      02-OCT-2000; 2000US-0679426.
PR      10-OCT-2000; 2000US-0685166.
XX
XX      (CORI-) CORIXA CORP.
XX
PI      Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI      Fanger GR, Reltter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI      Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;
XX      MPI; 2001-639232/73.
XX      N-PSDB; AAS64132.
DR
XX      New human prostate-specific polypeptides and polynucleotides useful for
PI      the diagnosis and treatment of cancer, especially prostate cancer -
XX
XX      Example 17: Page 533-534; 579pp; English.
PS
XX      The invention relates to isolated prostate-specific
CC      polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC      antibodies raised against the polypeptides (or antigenic epitopes
CC      derived from them) and antigen-presenting cells expressing the
CC      polypeptides. The antibodies are useful for detecting the presence of
CC      cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC      the antigen-presenting cells are useful for stimulating and/or expanding
CC      T cells specific for a tumour protein, and for inhibiting the development
CC      of cancer especially prostate cancer. Compositions comprising the
CC      polynucleotide and/or polypeptide are useful for stimulating an immune
CC      response, and for treating cancer. The oligonucleotide is useful for
CC      detecting cancer. The present sequence is fusion protein comprising a
CC      prostate specific polypeptide of the invention.
XX
SQ      Sequence 224 AA:

```

```

Query Match          100.0%; Score 148; DB 22; Length 224;
Best Local Similarity 100.0%; Pred. No. 2,4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 TAAADNFQLSQGGGQGFAPIGQAMAIACQI 30
      |||
      8 taasdnfqlsggggqgfapigqamataagqi 37

```

Search completed: August 13, 2002, 10:37:10
Job time: 119 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:32 ; Search time 12.82 Seconds
(without alignments)
57.158 Million cell updates/sec

Title: US-09-684-215A-17

148

Perfect score: 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30

Sequence: BLOSOM62

Gapop 10.0 , Gapext 0.5

Scoring table: 231628 seqs, 24425594 residues

Searched: Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 30

Maximum DB seq length: 30

Post-processing: Minimum Match 100%

Maximum Match 100%

Listing first 100 summaries

Database: Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/PCOMUS_COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	100.0	30	4	US-09-439-313-484

ALIGNMENTS

RESULT 1
US-09-439-313-484
Sequence 484, Application US/09439313
Patent No. 623505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yugu
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Saplen
US-09-439-313-484

Query Match 100.0%; Score 148; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30
|||||
Db 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30

Search completed: August 13, 2002, 10:58:53
Job time: 21 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:12 ; Search time 25.01 Seconds
(without alignments) 115.261 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148

Sequence: 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_71:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	355	2	F70983
2	104.5	70.6	361	2	S47170
3	98.5	66.6	354	2	A87242
4	96.5	66.6	354	2	F87383
5	95.5	66.6	354	2	C70821
6	94.5	66.6	354	2	H86930
7	93.5	66.6	354	2	AG2150
8	92.5	66.6	354	2	S75338
9	91.5	66.6	354	2	T45448
10	90.5	66.6	354	2	B54589
11	89.5	66.6	354	2	A12516
12	88.5	66.6	354	2	A86247
13	87.5	66.6	354	2	H90988
14	86.5	66.6	354	2	B85834
15	85.5	66.6	354	2	F86554
16	84.5	66.6	354	2	G72069
17	83.5	66.6	354	2	AD1894
18	82.5	66.6	354	2	C82630
19	81.5	66.6	354	2	T35287
20	80.5	66.6	354	2	A96739
21	79.5	66.6	354	2	S16306
22	78.5	66.6	354	2	F90707
23	77.5	66.6	354	2	B85558
24	76.5	66.6	354	2	G83658
25	75.5	66.6	354	2	G89509
26	74.5	66.6	354	2	G64573
27	73.5	66.6	354	2	C96753
28	72.5	66.6	354	2	S74643
29	71.5	66.6	354	2	F83472

30	49	33.1	665	2	AG1117	transketolase homo
31	49	33.1	665	2	A11477	transketolase homo
32	49	33.1	1194	2	D49851	magnesium-protopor
33	48	32.4	350	2	S34557	hypothetical prote
34	48	32.4	506	2	G86806	lipopolysaccharide
35	48	32.4	697	2	T34704	probable transfera
36	48	32.4	822	2	E75523	ATP-dependent heli
37	47.5	32.1	341	2	S33464	hypothetical prote
38	47.5	32.1	342	2	A96581	hypothetical prote
39	47.5	32.1	581	2	T51580	ovule development
40	47.5	32.1	664	2	A86296	hypothetical prote
41	47.5	32.1	700	2	C86296	hypothetical prote
42	47.5	32.1	1487	2	G96827	protein F20817.10
43	47	31.8	257	2	A72395	oxidoreductase, sh
44	47	31.8	351	2	S29979	hype protein - Alc
45	47	31.8	412	2	JC1116	type III site-spec

ALIGNMENTS

RESULT 1
F70983
Probable serine proteinase pepa - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Andrews, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID:98295987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-355 <CON>
A: Cross-references: GB:296071; GB:A123456; NID:93242254; PIDN:CAB09453.1; PID:g21819
A: Experimental source: strain H37RV
C: Geneticals:
A: Gene: pepa
C: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 100.0%; Score 148; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. NO. 1.2e-13;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30
DB 224 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 253

RESULT 2
S47170
hypothetical protein 34k - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34kDa protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-361 <CAN>
A: Cross-references: EMBL:Z23092; NID:9505550; PIDN:CAA80638.1; PID:9505551
C: Superfamily: proteinase h30b

Query Match 70.6%; Score 104.5; DB 2; Length 361;

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: AB6247
 R:Phologist: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719
 A:Accession: AB6247
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-820 <STO>
 A:Cross-references: GB:AE005172; NID:g5734727; PIDN:AAD4992.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein

Query Match 34.8%; Score 51.5; DB 2; Length 820;
 Best Local Similarity 45.2%; Pred. No. 24;
 Matches 14; Conservative 4; Mismatches 6; Indels 7; Gaps 2;

OY 3 ASDNQLS---QGQGFPIPIGQAMATAGQ 29
 Db 505 ATDNFSLSNKLGQGGFG--PYKGMLEGG 532

RESULT 13
 H90988
 hypothetical protein ECS2880 [imported] - Escherichia coli (strain O157:H7, substrain R1
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: H90988
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawaga, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: H90988
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <HAV>
 A:Cross-references: GB:BA000007; PIDN:BA836303.1; PID:g13362349; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: ECS2880
 C:Superfamily: Escherichia coli hypothetical protein b2072

Query Match 34.5%; Score 51; DB 2; Length 253;
 Best Local Similarity 47.6%; Pred. No. 8.2;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSQGGGFAIPIGQAMATAGQ 29
 Db 51 VSQGGGAMLAINEAMATVSO 71

RESULT 14
 B85834
 hypothetical protein Z3240 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B85834
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 525-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B85834
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <STO>
 A:Cross-references: GB:AE005174; NID:g12516274; PIDN:AAG57134.1; GSPDB:GN00145; UNCP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3240
 C:Superfamily: Escherichia coli hypothetical protein b2072

Query Match 34.5%; Score 51; DB 2; Length 253;
 Best Local Similarity 47.6%; Pred. No. 8.2;
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSQGGGFAIPIGQAMATAGQ 29
 Db 51 VSQGGGAMLAINEAMATVSO 71

RESULT 15
 E86554
 Fe-S oxidoreductase [imported] - Chlamydia pneumoniae (strain J138)
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: E86554
 R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; MUID:20330349
 A:Accession: E86554
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-369 <STO>
 A:Cross-references: GB:BA000008; NID:g8978883; PIDN:BA98719.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: CPj0513
 C:Superfamily: hypothetical protein AF0390

Query Match 34.5%; Score 51; DB 2; Length 369;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 8 QLSQGGGFAIPIGQAMATAGQ 27
 Db 272 RLNRSGGHAIPILKSLMAYVA 291

Search completed: August 13, 2002, 10:36:14
 Job time: 62 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:36:17 ; Search time 15.81 Seconds
(without alignments)
73.472 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148
Sequence: 1 TAAADFOLFQSGGCFAPFGQMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5.

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.5	36.1	239	1	PSB6_HUMAN
2	49.5	33.4	416	1	P24077 escherichia
3	49	33.1	253	1	YBCK_ECOLI
4	49	33.1	1194	1	BCBH_RHOC
5	48	32.4	350	1	YCKA_EUGR
6	47.5	32.1	694	1	ABC7_MOUSE
7	47	31.8	351	1	HYRE_ALCEU
8	47	31.8	412	1	T3RE_BACCE
9	46.5	31.4	225	1	Y130_MYCTU
10	46.5	31.4	752	1	ABC7_HUMAN
11	46.5	31.4	928	1	PMP9_CHLPN
12	46	31.1	1609	1	CYPI_MYCLE
13	45.5	30.7	376	1	CARA_MYCTU
14	45	30.4	396	1	REFX_SHIDY
15	45	30.4	541	1	60TM_HAEIN
16	44.5	30.1	237	1	PSB6_RAT
17	44.5	30.1	238	1	PSB6_MOUSE
18	44	29.7	288	1	Y012_MYCPN
19	44	29.7	351	1	MRAY_BORBU
20	44	29.7	382	1	AZAB_DIDMA
21	43.5	29.4	119	1	CHAI_ANTPO
22	43.5	29.4	776	1	VP4_ROTTH6
23	43.5	29.4	776	1	VP4_ROTTH6
24	43	29.1	233	1	ROCI_SPTOL
25	43	29.1	306	1	RBSK_HAEIN
26	43	29.1	432	1	Y1S4_YEAST
27	43	29.1	444	1	PSR1_YEAST
28	43	29.1	461	1	SG2N_RAT
29	43	29.1	628	1	SNXI_HUMAN
30	43	29.1	700	1	FLHA_CAUCR
31	43	29.1	1193	1	BCBH_RHOSH
32	43	29.1	1362	1	PMD1_SCHPO
33	43	29.1	1530	1	RPCI_TRYBB

34	42.5	28.7	110	1	RPO2_MYCLE
35	42.5	28.7	617	1	PSY1_PSEAE
36	42.5	28.7	688	1	PSY2_PSEAE
37	42.5	28.7	776	1	VP4_ROTBU
38	42.5	28.7	776	1	VP4_ROTBU
39	42.5	28.7	1752	1	RBI1_SCHRO
40	42	28.4	238	1	RISA_YEAST
41	42	28.4	323	1	CDIC2_EMENT
42	42	28.4	339	1	NRL2_ARATH
43	42	28.4	346	1	NRL1_ARATH
44	42	28.4	450	1	CRED_ECOLI
45	42	28.4	455	1	DEGO_ECOLI

ALIGNMENTS

RESULT 1	PSB6_HUMAN	STANDARD:	PRT: 239 AA.
ID	PSB6_HUMAN		
AC	P28072;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Proteasome subunit beta type 6 precursor (EC 3.4.25.1) (Proteasome delta chain) (Macropain delta chain) (Multicatalytic endopeptidase complex delta chain) (Proteasome subunit Y).		
GN	PSMB6 OR LMPY.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE-94345396; PubMed-8066462;		
RA	Akiyama K.-Y., Yokota K.-Y., Kagawa S., Shimbara N., Tamura T., Akioka H., Notwang H.G., Noda C., Tanaka K., Ichihara A.;		
RT	"CDNA cloning and interferon gamma down-regulation of proteasomal subunits X and Y."		
RL	Science 265:1231-1234(1994).		
RL	[2]		
RP	SEQUENCE OF 60-239 FROM N.A., AND SEQUENCE OF 35-75; 80-110 AND 210-233.		
RN	MEDLINE-91363412; PubMed-1888762;		
RA	Demartino G.N., Orth K., McCullough M.L., Lee L.W., Munn T.Z., Moomaw C.R., Dawson P.A., Slaughter C.A.;		
RT	"The primary structures of four subunits of the human, high-molecular-weight proteinase, macropain (proteasome), are distinct but homologous."		
RL	Biochim. Biophys. Acta 1079:29-38(1991).		
RP	SEQUENCE OF 35-60.		
RX	MEDLINE-90167111; PubMed-2306472;		
RA	Lee L.W., Moomaw C.R., Orth K., McGuire M.J., Demartino G.N., Slaughter C.A.;		
RT	"Relationships among the subunits of the high molecular weight proteinase, macropain (proteasome)."		
RL	Biochim. Biophys. Acta 1037:178-185(1990).		
CC	-I- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC ACTIVITY. MAY CATALYZE BASAL PROCESSING OF INTRACELLULAR ANTIGENS.		
CC	-I- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad specificity.		
CC	-I- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal proteolytic pathway.		
CC	-I- SUBUNIT: The proteasome is composed of at least 15 non identical subunits which form a highly ordered ring-shaped structure. This subunit can be displaced by the equivalent immune-specific subunit PSMB9.		
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.		
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B; ALSO KNOWN AS THE		

```

CC      PROTEASOME B-TYPE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D29012; BA006098.1; -
DR      EMBL; X61971; CAA43963.1; -
DR      PIR; S08188; S08188.
DR      PIR; S17522; S17522.
DR      MEROPS; T01.010; -
DR      SWISS-2DPAGE; P28072; HUMAN.
DR      PHCI-2DPAGE; P28072; -
DR      MIM; 600307; -
DR      InterPro; IPR001353; Proteasome.
DR      InterPro; IPR000243; Proteasome_B.
DR      Pfam; PF00227; Proteasome; 1.
DR      PRINTS; PR00141; PROTEASOME.
DR      PROSITE; PS00854; PROTEASOME_B; 1.
KW      Proteasome; Hydrolase; Protease; Zymogen.
FT      PROPEP 1 34
FT      CHAIN 35 239
FT      CONFLICT 145 145 G -> V (IN REF. 2).
SQ      SEQUENCE 239 AA; 25315 MW; 66EBB9B6C685830D CRC64;

Query Match 36.1%; Score 53.5; DB 1; Length 239;
Best Local Similarity 43.5%; Pred. No. 1.7;
Matches 10; Conservative 6; Mismatches 2; Indels 5; Gaps 1;

QY 11 QGGGFAPIG-----QMAING 28
Db 141 EGGGYSYPMGMMVRSFAIGG 163

RESULT 2
YBDA_ECOLI
ID YBDA_ECOLI STANDARD; PRT; 416 AA.
AC P24077;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical membrane protein P43.
GN YBDA OR B0591.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92157868; PubMed-1838574;
RA Shea C.M., McIntosh M.A.;
RT "Nucleotide sequence and genetic organization of the ferric
RT enterobactin transport system: homology to other periplasmic binding
RT protein-dependent systems in Escherichia coli.";
RL Mol. Microbiol. 5:1415-1428(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92157867; PubMed-1787794;
RA Chenault S.S., Earhart C.F.;
RT "Organization of genes encoding membrane proteins of the Escherichia
RT coli ferrienterobactin permease.";
RL Mol. Microbiol. 5:1405-1413(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed-9278503;

```

```

RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federapfel N., Hyman R., Kaiman S., Komp C., Kardi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X57470; CAA40706.1; -
DR      EMBL; X59402; CAA42044.1; -
DR      EMBL; AE000164; AAC73692.1; -
DR      EMBL; U82598; ABA40790.1; -
DR      PIR; S16295; S16295.
DR      PIR; S16306; S16306.
DR      PIR; S14850; S14850.
DR      EcoGene; EG11104; ybda.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT FT TRANSMEM 22 42
FT FT TRANSMEM 56 76
FT FT TRANSMEM 96 116
FT FT TRANSMEM 119 139
FT FT TRANSMEM 157 177
FT FT TRANSMEM 179 199
FT FT TRANSMEM 219 239
FT FT TRANSMEM 257 277
FT FT TRANSMEM 301 321
FT FT TRANSMEM 357 377
FT FT TRANSMEM 379 399
FT FT CONFLICT 123 123 F -> S (IN REF. 1).
SQ SEQUENCE 416 AA; 43282 MW; 64FCCF9AF99AC25A CRC64;

Query Match 33.4%; Score 49.5; DB 1; Length 416;
Best Local Similarity 46.2%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 6; Indels 3; Gaps 1;

QY 2 AASDNFQSGGGG--AIPGGAM 24
Db 248 ALADNMQMSAAGFLYAIPLGAI 273

RESULT 3
YEGK_ECOLI
ID YEGK_ECOLI STANDARD; PRT; 253 AA.
AC P76395;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yegK.
GN YEGK OR B2072.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

```


RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000297; AAC75133.1; --
 DR Ecogene: BG14054; yegK.
 DR Hypothetical protein: Complete proteome.
 KM SEQUENCE 253 AA; 27208 MW; DC828DD5AD67E7E0 CRC64;
 SO

Query Match 33.1%; Score 49; DB 1; Length 253;
 Best Local Similarity 42.9%; Pred. No. 7.6;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 9 LSGGCGFAIPIGAMATNGO 29
 DB 51 VSGGEGAMLAIVEMAYMSQ 71

RESULT 4
 BGHH_RHOCA STANDARD; PRT; 1194 AA.
 ID BGHH_RHOCA
 AC P26162;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Magnesium-chelatase subunit H (Mg-protoporphylin IX chelatase subunit
 DE H).
 GN BGHH.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxID=1061;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003;
 RX MEDLINE=93224465; PubMed=8385667;
 RA Burke D.H., Alberti M., Hearst J.E.;
 RA "bchNbh bacteriochlorophyll synthesis genes of Rhodobacter
 RT capsulatus and identification of the third subunit of
 RT light-independent protochlorophyllide reductase in bacteria and
 RT plants.";
 RT J. Bacteriol. 175:2414-2422(1993).
 RL J. Bacteriol. 175:5001-5010(1990).
 RT bacteriochlorophyll biosynthetic pathway.";
 RA "Rhodobacter capsulatus genes involved in early steps of the
 RT bacteriochlorophyll biosynthetic pathway.";
 RL J. Bacteriol. 172:5001-5010(1990).
 CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
 CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
 CC PROTOPORPHYRIN IX.
 CC -1- PATHWAY: LIGHT-INDEPENDENT BACTERIOCHLOROPHYLL BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNIT H FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

CC EMBL: 211165; AAA7524.1; --
 DR EMBL: M34843; AAA26097.1; --
 DR PIR: S17808; S17808.
 DR PIR: A36716; A36716.
 DR PIR: D49851; D49851.
 DR InterPro: IPR003672; COBN/Mg-chelatase.
 DR Pfam: PF02514; COBN-Mg_chel; 1.
 KW Photosynthesis; Bacteriochlorophyll biosynthesis.
 KM SEQUENCE 1194 AA; 129361 MW; 6341816A58774EE5 CRC64;
 SO

Query Match 33.1%; Score 49; DB 1; Length 1194;
 Best Local Similarity 52.0%; Pred. No. 35;
 Matches 13; Conservative 3; Mismatches 3; Indels 6; Gaps 2;

OY 4 SDNFQSGGCGFAIPIGAMATNG 28
 DB 848 SDNIR-SDG-----PIGQALALMG 866

RESULT 5
 YCXA_EUGCR STANDARD; PRT; 350 AA.
 ID YCXA_EUGCR
 AC P31561;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 40.9 kDa protein in 16S rRNA 3' region (ORF350).
 DE Euglena gracilis.
 OS Chloroplast.
 OG Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OX NCBI_TaxID=3039;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RX MEDLINE=93347989; PubMed=8346031;
 RA Halliack R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,
 RA Orsat B., Spielmann A., Stutz E.;
 RA "Complete sequence of *Euglena gracilis* chloroplast DNA.";
 RL Nucleic Acids Res. 21:3537-3544(1993).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RA Schluenger B., Stutz E.;
 RA "The *Euglena gracilis* chloroplast genome: structural features of a
 RT DNA region possibly carrying the single origin of DNA replication.";
 RT Curr. Genet. 8:629-634(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 211874; -- NOT_ANNOTATED_CDS.
 DR EMBL: X70810; CAAS0140.1; --
 DR PIR: S34557; S34557.
 DR PIR: S36933; S36933.
 KW Chloroplast; Hypothetical protein.
 KM SEQUENCE 350 AA; 40861 MW; 1AACAT7EC540FF51 CRC64;
 SO

Query Match 32.4%; Score 48; DB 1; Length 350;
 Best Local Similarity 56.2%; Pred. No. 14;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 15 GFATPIGAMATNGOI 30
 DB 315 GFVYPGQMSAPGOI 330

```

RESULT 6
ID ABC_MOUSE STANDARD: PRT; 694 AA.
AC 061102;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family B, member 7, mitochondrial (ATP-binding cassette transporter 7) (ABC transporter 7 protein) (Fragment).
DE (Fragment).
GN ABC7 OR ABC7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DRB/2;
RA MEDLINE=97288528; PubMed=9143506;
RA Savary S., Allikmets R., Denizot F., Luciani M.-F., Mattei M.-G., Dean M., Chimi G.;
RT Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human.;
RL Genomics 41:275-278(1997).
CC -1- FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE MITOCHONDRIA TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial inner membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43892; AAC53152.1; -.
DR MGD: MGI:109333; Abcd7.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmem.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane.1.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane.
FT NON_TER 1 1
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT NP_BIND 447 454 ATP (POTENTIAL).
SQ SEQUENCE 694 AA; 76417 MW; A7AE89AE9AA9BID CRC64;

```

```

Query Match 32.1%; Score 47.5; DB 1; Length 694;
Best Local Similarity 37.1%; Pred. No. 33;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

```

```

QY 1 TASDN--POLSGG-----GQGAIRIGQAMAIAG 28
DB 413 TVAFDNVHEFYIEGQVYLVNGSVFEPAGKVAIVG 447

```

```

RESULT 7
ID HYPE_ALCEU STANDARD: PRT; 351 AA.

```

```

AC P31905;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1993 (Rel. 38, Last annotation update)
DE Hydrogenase expression/formation protein hype.
GN HYPE.
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Bacteri; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H16 / ATCC 17699;
RA MEDLINE=93356597; PubMed=8352644;
RA Dierckx J., Eitlinger M., Friedrich B.;
RT "Analysis of a pleiotropic gene region involved in formation of catalytically active hydrogenases in Alcaligenes eutrophus H16.";
RL Arch. Microbiol. 159:545-553(1993).
CC -1- SIMILARITY: BELONGS TO THE HYPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70183; CAA49735.1; -.
DR PIR: S29979; S29979.
DR InterPro: IPR000728; AIRS-related.
DR Pfam: PF00586; AIRS; 1.
DR Pfam: PF02769; AIRS_C; 1.
KW Plasmid.
SQ SEQUENCE 351 AA; 36596 MW; 35D7291974B7BBBD CRC64;

```

```

Query Match 31.8%; Score 47; DB 1; Length 351;
Best Local Similarity 46.4%; Pred. No. 20;
Matches 13; Conservative 2; Mismatches 11; Indels 2; Gaps 1;

```

```

QY 2 AASDNFQLSGGG--GRAIRIGQAMAI 27
DB 41 AAFDNWELRQNDQAAFAFAMPAGARVMA 68

```

```

RESULT 8
ID T3RE_BACCE STANDARD: PRT; 412 AA.
AC P25241;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type III restriction-modification system BceII0987IP enzyme res (EC 3.1.21.5) (Fragment).
GN RES.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10987;
RA MEDLINE=92267377; PubMed=1587478;
RA Hegna I.K., Karlstrom E.S., Lopez R., Kristensen T., Kolstoe A.-B.;
RT "A type-III DNA restriction and modification system in Bacillus cereus?";
RL Gene 114:149-150(1992).
CC -1- FUNCTION: THIS PROTEIN CUT THE DNA OUTSIDE OF THE RECOGNITION SITE. MAY ALSO ACT AS A HELICASE INVOLVED IN UNWINDING DNA AT THE CLEAVAGE SITE. PROTEIN ONLY REQUIRED FOR RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME (BY SIMILARITY).
CC

```


RA Bishop D.F.;
 RT "Human ABC7 transporter: gene structure and mutation causing x-linked
 RT sideroblastic anemia with ataxia with disruption of cytosolic
 RT iron-sulfur protein maturation.";
 RL Blood 96:3256-3264(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 5-752 FROM N.A.
 RX MEDLINE=99098366; PubMed=9883897;
 RA Csere P., Hill R., Kispal G.;
 RT "Identification of a human mitochondrial ABC transporter, the
 RT functional orthologue of yeast Atmtp.";
 RL FEBS Lett. 441:266-270(1998).
 CC -! FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF HEME FROM THE
 CC MITOCHONDRION TO THE CYTOSOL. PLAYS A CENTRAL ROLE IN THE
 CC MATURATION OF CYTOSOLIC IRON-SULFUR (FE/S) CLUSTER-CONTAINING
 CC PROTEINS.
 CC -! SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
 CC -! SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (potential).
 CC -! DISEASE: DEFECTS IN ABC7 ARE THE CAUSE OF X-LINKED SIDEROBLASTIC
 CC ANEMIA WITH ATAXIA (ASAT). ASAT IS A RECESSIVE DISORDER
 CC CHARACTERIZED BY AN INFANTILE TO EARLY CHILDHOOD ONSET OF
 CC NONPROGRESSIVE CEREBELLAR ATAXIA AND MILD ANEMIA WITH HYPOCHROMIA
 CC AND MICROCYTOSIS.
 CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB005289; BA28861.1; -;
 DR EMBL: AF038950; AAC39865.1; -;
 DR EMBL: AF133659; AAD33045.1; -;
 DR EMBL: AF241887; AAK20173.1; -;
 DR EMBL: AF241872; AAK20173.1; JOINED.
 DR EMBL: AF241873; AAK20173.1; JOINED.
 DR EMBL: AF241874; AAK20173.1; JOINED.
 DR EMBL: AF241875; AAK20173.1; JOINED.
 DR EMBL: AF241876; AAK20173.1; JOINED.
 DR EMBL: AF241877; AAK20173.1; JOINED.
 DR EMBL: AF241878; AAK20173.1; JOINED.
 DR EMBL: AF241879; AAK20173.1; JOINED.
 DR EMBL: AF241880; AAK20173.1; JOINED.
 DR EMBL: AF241881; AAK20173.1; JOINED.
 DR EMBL: AF241882; AAK20173.1; JOINED.
 DR EMBL: AF241883; AAK20173.1; JOINED.
 DR EMBL: AF241884; AAK20173.1; JOINED.
 DR EMBL: AF241885; AAK20173.1; JOINED.
 DR EMBL: AF241886; AAK20173.1; JOINED.
 DR EMBL: BC006323; AAH06323.1; -;
 DR EMBL: AF078777; AAD47141.1; -;
 DR MIM: 300135; -;
 DR MIM: 301310; -;
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tmam.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00064; ABC_membrane; 1.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane;
 KW Transic peptide; Disease mutation.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).

FT	CHAIN	?	752	ATP-BINDING CASSETTE, SUB-FAMILY B, MEMBER 7. POTENTIAL.
FT	TRANSMEM	260	280	POTENTIAL.
FT	TRANSMEM	291	311	POTENTIAL.
FT	TRANSMEM	383	403	POTENTIAL.
FT	NP_BIND	505	512	ATP (POTENTIAL).
FT	VARIANT	400	400	I -> M (IN ASAT).
FT	VARIANT	400	400	/FTID=VAR_009156.
FT	VARIANT	433	433	E -> K (IN ASAT; IMPAIRED MATURATION OF CYTOSOLIC FE/S PROTEINS).
FT	VARIANT	433	433	/FTID=VAR_012640.
FT	CONFLICT	56	56	O -> OQ (IN REF. 5).
FT	CONFLICT	141	141	A -> P (IN REF. 2).
FT	CONFLICT	258	258	R -> K (IN REF. 1).
FT	CONFLICT	271	276	LLPINE -> PLPNV (IN REF. 2).
FT	CONFLICT	281	282	VS -> LVS (IN REF. 2).
FT	CONFLICT	290	290	G -> C (IN REF. 2).
FT	CONFLICT	293	297	FALVT -> LLGN (IN REF. 2).
FT	CONFLICT	315	315	R -> G (IN REF. 1).
FT	CONFLICT	320	324	TEMNK -> LEIDQ (IN REF. 2).
FT	CONFLICT	346	346	F -> I (IN REF. 1).
FT	CONFLICT	542	542	E -> V (IN REF. 6).
FT	CONFLICT	542	542	E -> V (IN REF. 6).
SO	SEQUENCE	752 AA;	82641 MW;	BLF5A57ABD24FB90 CRC64;

Query Match 31.4%; Score 46.5; DB 1; Length 752;
 Best Local Similarity 37.1%; Pred.No. 50;
 Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

QY 1 TAASDN--FOLSOG-----GQGFAPIGQMAIAG 28
 Db 471 TVAFDNVHFYEYEGKVLGISFEVPARKKVAIVG 505

RESULT 11
 PMP9.CHLPN
 ID PMP9.CHLPN STANDARD; PRT; 928 AA.
 AC Q9Z398;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable outer membrane protein pmp9 precursor (Polymorphic membrane protein 9) (outer membrane protein 10).
 GN PMP9 OR OMP10 OR CPN0447 OR CP0306.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VRI310;
 RX MEDLINE=20007584; PubMed=10539856;
 RA Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygland P.,
 RA Madsen A.S., Knudsen K., Falk E., Birkedal S.;
 RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
 RT role in immunopathogenicity.";
 RL Am. Heart J. 136:S491-S495(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger U., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uitterlind T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;

```
RT      *Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT      pneumoniae AR39."
RL      Nucleic Acids Res. 28:1397-1406(2000).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN-J138;
RX      MEDLINE=20330349; PubMed=10871362;
RA      Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      *Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA."
RL      Nucleic Acids Res. 28:2311-2314(2000).
CC      -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC      (POTENTIAL).
CC      -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: AJ13034; CAB37069.1; -
DR      EMBL: AE001628; AAD18591.1; -
DR      EMBL: AE002192; AAF38163.1; -
DR      EMBL: AP002546; BAA98655.1; -
DR      TIGR: CP0306; -
DR      InterPro: IPR003368; DUF145.
DR      InterPro: IPR003357; OMP.
DR      Pfam: PF02415; DUF145; 1.
DR      Pfam: PF02385; OMP; 1.
KM      Outer membrane; Signal; Multigene family; Complete proteome.
FT      SIGNAL 1 26
FT      CHAIN 27 928
FT      SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;
SQ
Query Match 31.4%; Score 46.5; DB 1; Length 928;
Best Local Similarity 39.4%; Pred. No. 62;
Matches 13; Conservative 3; Mismatches 8; Indels 9; Gaps 1;
QY      4 SDNFOLSGGGGF-----AIPIGQMAIA 27
DB      293 TDNLVSSGGPTLFKNNSAIDTAAPLGAIAIA 325
RESULT 12
CTPI_MYCLE STANDARD; PRT; 1609 AA.
AC      053114;
DT      16-OCT-2001 (Rel. 40; Created)
DT      16-OCT-2001 (Rel. 40; Last sequence update)
DT      16-OCT-2001 (Rel. 40; Last annotation update)
DE      Probable cation-transporting ATPase I (EC 3.6.3.-).
GN      CTPI OR MD2671 OR MCB1913.02.
OS      Mycobacterium leprae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1769;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-TN;
RX      MEDLINE=21128732; PubMed=11234002;
RA      Cole S.T., Egidmeier K., Parkhill J., James K.D., Thomson N.R.,
RA      Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA      Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA      Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
RA      Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA      Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA      Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
```

```
RA      Barrell B.G.;
RT      *Massive gene decay in the leprosy bacillus."
RL      Nature 409:1007-1011(2001).
CC      -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC      (E1-E2 ATPASES).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: AL022118; CAAT17934.1; -
DR      EMBL: AL583926; CAC32203.1; -
DR      Leptoma; M12671; -
DR      InterPro: IPR001757; E1-E2_ATPase.
DR      InterPro: IPR001454; Hydrolase.
DR      Pfam: PF00122; E1-E2_ATPase; 1.
DR      Pfam: PF00702; Hydrolase; 1.
DR      PRINTS: PR00119; CATATPASE.
DR      PROSITE: PS00154; ATPASE_E1_E2; 1.
KM      Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW      Complete proteome.
FT      TRANSMEM 30 50
FT      TRANSMEM 176 196
FT      TRANSMEM 238 258
FT      TRANSMEM 357 377
FT      TRANSMEM 641 661
FT      TRANSMEM 673 693
FT      TRANSMEM 778 798
FT      TRANSMEM 921 941
FT      TRANSMEM 969 989
FT      TRANSMEM 997 1017
FT      TRANSMEM 1396 1416
FT      TRANSMEM 1426 1446
FT      TRANSMEM 1542 1562
FT      TRANSMEM 1573 1593
FT      MOD_RES 1053 1053
FT      METAL 1335 1335
FT      METAL 1339 1339
FT      SEQUENCE 1609 AA; 166871 MW; 58FA2079905E3995 CRC64;
SQ
Query Match 31.1%; Score 46; DB 1; Length 1609;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY      12 GGGGFAIPIGQMAIAGOI 30
DB      388 GQQAFAATLGRGLANAGOL 406
RESULT 13
CARA_MYCTU STANDARD; PRT; 376 AA.
AC      P71811;
DT      16-OCT-2001 (Rel. 40; Created)
DT      16-OCT-2001 (Rel. 40; Last sequence update)
DT      16-OCT-2001 (Rel. 40; Last annotation update)
DE      Carbamoyl-phosphate synthase small chain (EC 6.3.5.5) (Carbamoyl-
DE      phosphate synthetase glutamine chain).
GN      CARA OR RV1383 OR MT1427 OR MTCY02B12.17.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-H37RV;
```

Query Match	30.7%;	Score 45.5;	DB 1;	Length 376;
Best Local Similarity	44.0%;	Pred. No. 35;		
Matches	11;	Conservative	3;	Mismatches 10;
				Indels 1;
				Gaps 1;
OY	1	TAASDNFOL-SQGQGFAPIGQAM	24	
	11	: 1 1 : 1 1 1 1 1 :		
Db	299	TAONHGFAILOGEAGOSFATPGPAV	323	
RESULT' 14				
PFBX_SHIDY		STANDARD;	PRT;	396 AA.
ID	REFBX_SHIDY			

Query Match	30.4%	Score 45;	DB 1;	Length 396;
Best Local Similarity	52.9%	Pred. NO. 43;		
Matches	9;	Conservative	2;	Mismatches
			6;	Indels
			0;	Gaps
			0;	
Qy	8	QLSGGGGFAIPIGAM	24	
	:	11	:	
Db	246	KLRMAAGLLIPIGAV	262	
RESULT	15			
60IM_HAEIN				
ID	60IM_HAEIN	STANDARD;	PRT;	541 AA.
AC	P44973;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	60 kDa inner-membrane protein homolog.			
GN	H11001.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision: Pasteurellales; Pasteurellales; Pasteurellaceae; Bacteria; Proteobacteria; gamma subdivision: Pasteurellales; Pasteurellales; Pasteurellaceae;			

```

OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Giodè A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.":
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE OXAL / 60 KDA IMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32781; AAC22663.1; -.
DR TIGR: H11001; -.
DR InterPro: IPR001708; 60KD_inner_MP.
DR Pfam: PF02096; 60KD_IMP_1.
DR PRINTS: PR00701; 60KDINNERMP.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 495 515 POTENTIAL.
SQ SEQUENCE 541 AA; 61239 MW; 9406398B334FAEDE CRC64;

```

```

Query Match 30.4%; Score 45; DB 1; Length 541;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
QY 5 DNFOLSGGGGFAIP 20
DB 139 DNFKLAEGGESLSVPL 154

```

Search completed: August 13, 2002, 10:42:45
Job time: 388 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:47 ; Search time 43.59 Seconds
(without alignments)
119.061 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAASDNFQLSGGCGFAIPGMAIAIGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	355	16	007175 mycobacteri
2	104.5	70.6	361	2	050320 mycobacteri
3	98.5	66.6	354	16	09CCY9 mycobacteri
5	57	38.5	780	4	096C66 homo sapien
5	56	37.8	296	16	09A982 caulobacter
6	55	37.2	464	16	053886 mycobacteri
7	55	37.2	542	2	09FBR9 streptomyc
8	54	36.5	125	4	09BVI4 homo sapien
9	54	36.5	382	16	09CD67 mycobacteri
10	54	36.5	407	4	096K50 homo sapien
11	54	36.5	452	2	092566 mycobacteri
12	54	36.5	452	16	093354 mycobacteri
13	53	35.8	419	11	099M10 mus musculu
14	53	35.8	419	11	091YE2 mus musculu
15	52	35.1	519	2	093J30 streptomyc
16	52	35.1	630	11	099JIV6 mus musculu

17	51.5	34.8	820	10	095XB5	095XB5 arbidopsi
18	51	34.5	369	16	092B39	092B39 chlamydia p
19	50.5	34.1	1000	16	09PCD0	09PCD0 xylella fas
20	50	33.8	362	2	09S2R5	09S2R5 streptomyc
21	50	33.8	487	10	09LPN6	09LPN6 arbidopsi
22	49.5	33.4	358	10	096520	096520 arbidopsi
23	49.5	33.4	358	10	043734	043734 arbidopsi
24	49.5	33.4	353	16	09K6J0	09K6J0 bacillus ha
25	49.5	33.4	571	2	09L4P0	09L4P0 staphylococ
26	49.5	33.4	662	16	099UD4	099UD4 staphylococ
27	49.5	33.4	784	12	091B33	091B33 equine rhin
28	49	33.1	275	10	09LIX7	09LIX7 oryza sativ
29	49	33.1	340	10	023632	023632 arbidopsi
30	49	33.1	340	10	023632	023632 arbidopsi
31	49	33.1	394	16	P72780	P72780 synchocyst
32	49	33.1	591	16	0913V3	0913V3 pseudomonas
33	49	33.1	665	16	092E08	092E08 listeria in
34	49	33.1	996	5	095XN3	095XN3 caenorhabdi
35	49	33.1	2553	5	022860	022860 caenorhabdi
36	48	32.4	348	16	0926M0	0926M0 listeria in
37	48	32.4	374	2	08S519	08S519 escherichia
38	48	32.4	374	2	093Q24	093Q24 salmonella
39	48	32.4	407	16	098J66	098J66 rhizobium 1
40	48	32.4	473	2	044476	044476 azotobacter
41	48	32.4	481	2	09AOD1	09AOD1 pseudomonas
42	48	32.4	506	16	09CFI4	09CFI4 lactococcus
43	48	32.4	697	2	069854	069854 streptococ
44	48	32.4	822	16	09RX95	09RX95 delinococcus
45	47.5	32.1	341	10	093104	093104 arbidopsi

ALIGNMENTS

RESULT 1
ID 007175 PRELIMINARY; PRT; 355 AA.
AC 007175;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE HYPOTHEICAL 34.9 KDA PROTEIN.
GN PEPA OR RV0125 OR MTC1418B.07.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -! SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: Z96071; CAB09453.1; -
DR MEROPS: S01.UPC;
DR Tuberculin: RV0125; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ. 1.
DR Pfam: PF00089; trypsin. 1.
DR SMART: SM00226; PDZ. 1.
DR PROSITE: PS50106; PDZ. 1.
DR PROSITE: PS50240; TRYPSIN_DOM. 1.

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN.L.
KW Complete proteome: Hydrolase; Hypothetical protein: Serine protease.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match
Best Local Similarity 100.0%; Score 148; DB 16; Length 355;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
Db 224 TAASNFOLSGGGGFAIPIGQAMAIAGQI 253

RESULT 2
ID 050320 PRELIMINARY; PRT; 361 AA.
AC 050320;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 34KDA PROTEIN PRECURSOR
OS Mycobacterium paratuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD88/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameleon R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterization of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
DR EMBL: Z23092; CAAB0638.1; -
DR MEROPS: S01.UPC; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 38 POTENTIAL.
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match
Best Local Similarity 70.6%; Score 104.5; DB 2; Length 361;
Matches 21; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
Db 231 TAATDSYKMS-GGOGFAIPIGRAMAVANQI 259

RESULT 3
ID 09CCY9 PRELIMINARY; PRT; 354 AA.
AC 09CCY9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROBABLE SECRETED SERINE PROTEASE.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;

RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean A., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL583926; CAC32191.1; -
DR MEROPS: S01.UPC; -
DR Leproma; ML2659; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR000126; Ser-proteas_V8.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A4 CRC64;

Query Match
Best Local Similarity 66.7%; Score 98.5; DB 16; Length 354;
Matches 20; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 1 TAASNFOLSGGGGFAIPIGQAMAIAGQI 30
Db 224 TAATDNYKM-LGGGFAIPIGQAMEYVGA 252

RESULT 4
ID 096C66 PRELIMINARY; PRT; 780 AA.
AC 096C66;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO RIKEN CDNA 1810073N04 GENE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC014642; AAH14642.1; -
SQ SEQUENCE 780 AA; 83324 MW; E529A228C842312 CRC64;

Query Match
Best Local Similarity 38.5%; Score 57; DB 4; Length 780;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 3 ASDNFOLSGGGGFAIPITGQA 23
Db 474 APEAQLDRPGGIAVPVGEA 494

RESULT 5
ID 09A9B2 PRELIMINARY; PRT; 296 AA.

```

AC 09A982:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC1082.
GN CC1082.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OC NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernm W.C., Feldhlym T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Kolonay J.F., Smt J., Durkin A.S., Gwinn M.L., Hatt D.H.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005785; AAK23066.1; -.
DR TIGR; CC1082; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 296 AA; 31983 MW; 2EB4EBA376823924 CRC64;

Query Match 37.8%; Score 56; DB 16; Length 296;
Best Local Similarity 37.5%; Pred. No. 5.1;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 4 SDNFOLSGGGFAIPGQAMAIYA 27
| : : | : : | : : | : : |
Db 159 SVEFEIMTGSVMVPGVAIAVA 182

RESULT 6
053896 PRELIMINARY; PRT; 464 AA.
AC 053896:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE SERINE PROTEASE.
GN RV0983 OR MTW044.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares K.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL021999; CAAL1582.1; -.
DR MEROPS; S01.0PC; -.
DR TubercuList; RV0983; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR HydroLase; Protease; Serine protease.
KW SEQUENCE 542 AA; 53761 MW; 620F7D889DE7212D CRC64;

```

```

DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; HydroLase; Protease; Serine protease.
SQ SEQUENCE 464 AA; 46452 MW; AE3BRC53E1BC8F CRC64;

Query Match 37.2%; Score 55; DB 16; Length 464;
Best Local Similarity 40.6%; Pred. No. 12;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TASNDFOLSGG--GGFAIPGQAMAIAGOI 30
| : : | : : | : : | : : |
Db 337 TLGADSDAAGSGSIGGFAIPVDQAKRIADEL 368

RESULT 7
09FBK9 PRELIMINARY; PRT; 542 AA.
ID 09FBK9
AC 09FBK9:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE PROTEASE.
GN SCP8.12.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinshah H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 mb streptomyces coelicolor A3(2) chromosome.";
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL390975; CAC01350.1; -.
DR MEROPS; S01.0PC; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR01217; FRICHEXTENS.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW HydroLase; Protease; Serine protease.
SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DE7212D CRC64;

Query Match 37.2%; Score 55; DB 2; Length 542;
Best Local Similarity 44.4%; Pred. No. 14;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 4 SDNFOLSGGGFAIPGQAMAIAGOI 30
| : : | : : | : : | : : |
Db 412 SDDGAGSIGGFAIPINQGRVAEEL 438

```


DR MEROPS; S01.UPC; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR HydroLase: Protease; Serine protease.
KW SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 36.5%; Score 54; DB 2; Length 452;
Best Local Similarity 44.8%; Pred. No. 16;
Matches 13; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 2 AASDNFOLSGGGGFAIPGAMAIAGOI 30
DB 328 ADSGDAQSGSIGLGFALPVDQAKRIADEL 356

RESULT 12
P73354 PRELIMINARY; PRT; 452 AA.
AC P73354;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE SERINE PROTEASE HTRA.
GN HTRA OR SLR1204.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: D90905; BAA17385.1; -.
DR MEROPS; S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; Trypsin; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR Complete proteome: HydroLase; Protease; Serine protease.
KW SEQUENCE 452 AA; 47656 MW; AED4AFB1AC47361 CRC64;

Query Match 36.5%; Score 54; DB 16; Length 452;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 15; Conservative 2; Mismatches 9; Indels 4; Gaps 1;

QY 1 TAASDNFOLSGGGGFAIPGAMAIAGOI 30
DB 312 TAIQNAQ-----GICFALPINKAQETAIQOI 337

RESULT 13
Q99MI0 PRELIMINARY; PRT; 419 AA.
AC Q99MI0;

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CELL GROWTH REGULATOR FALKOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C.
RA Erez N., Gudkov A.V., Peles E., Cohen I.R., Rotter V.;
RT "Falkor is a novel cell growth regulator, isolated from a MEF-derived
RT GSE library."
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF340231; AAK37525.1; -.
KW SEQUENCE 419 AA; 45108 MW; 56891872CAFBAB2B9 CRC64;

Query Match 35.8%; Score 53; DB 11; Length 419;
Best Local Similarity 42.9%; Pred. No. 20;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 AASDNFOLSGGGGFAIPGQ 22
DB 394 AARDKYOLASGQKGVQVPSQ 414

RESULT 14
Q91YE2 PRELIMINARY; PRT; 419 AA.
AC Q91YE2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE EGLN2 PROTEIN.
GN EGLN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21458562; PubMed=11574160;
RA Taylor M.S.;
RT "Characterisation and comparative analysis of the EGLN gene family.";
RL Gene 275:125-132(2001).
DR EMBL: AJ310547; GAC42516.1; -.
KW SEQUENCE 419 AA; 44937 MW; ED9A600BBB5886BF CRC64;

Query Match 35.8%; Score 53; DB 11; Length 419;
Best Local Similarity 42.9%; Pred. No. 20;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 AASDNFOLSGGGGFAIPGQ 22
DB 394 AARDKYOLASGQKGVQVPSQ 414

RESULT 15
Q93J30 PRELIMINARY; PRT; 519 AA.
AC Q93J30;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE PROTEASE.
GN SCBAC25E3.14.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:32 ; Search time 28.73 seconds
(without alignments)
115.984 Million cell updates/sec

Title: US-09-684-215a-17

Perfect score: 148
Sequence: 1 TAASDNFQLSGGGGFAIPGQAMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	30	22	AAU69788
2	148	100.0	30	22	AAU01143
3	148	100.0	30	22	AAU69028

ALIGNMENTS

RESULT 1
AAU69788
ID AAU69788 standard; Peptide: 30 AA.

XX AC AAU69788;
XX 30-JAN-2002 (first entry)
XX
XX
XX DE M. tuberculosis antigen Ra12 amino acids 1-30.
XX
XX KW Human; prostate cancer; ss; cytosolic; immunostimulant; tumour;
XX antigen; epitope.
XX OS Mycobacterium tuberculosis.
XX
XX PN W0200173032-A2.
XX PD 04-OCT-2001.
XX PF 27-MAR-2001; 2001WO-US09919.
XX PR 27-MAR-2000; 2000US-0536857.
XX PR 09-MAY-2000; 2000US-0568100.
XX PR 12-MAY-2000; 2000US-0570737.
XX PR 13-JUN-2000; 2000US-0593793.
XX PR 27-JUN-2000; 2000US-0605783.
XX PR 10-AUG-2000; 2000US-0636215.
XX PR 29-AUG-2000; 2000US-0651236.
XX PR 06-SEP-2000; 2000US-0657279.
XX PR 02-OCT-2000; 2000US-0679426.
XX PR 10-OCT-2000; 2000US-0685166.
XX
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
XX Fanger GR, Retter MM, Stolk JA, Day CH, Vedvick TS, Carter D;
XX PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
XX
XX DR WPI; 2001-639232/73.
XX
XX PT New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer -
XX
XX PS Example 17; Page 395; 579pp; English.
XX
XX CC The invention relates to isolated prostate-specific
XX polynucleotides, polypeptides, fusion proteins of the polypeptides,
XX antibodies raised against the polypeptides (or antigenic epitopes
XX derived from them) and antigen-presenting cells expressing the
XX polypeptides. The antibodies are useful for detecting the presence of
XX cancer, especially prostate cancer. The polypeptides, polynucleotides and
XX the antigen-presenting cells are useful for stimulating and/or expanding
XX T cells specific for a tumour protein, and for inhibiting the development
XX of cancer especially prostate cancer. Compositions comprising the
XX polynucleotide and/or polypeptide are useful for stimulating an immune
XX response, and for treating cancer. The oligonucleotide is useful for
XX detecting cancer. The present sequence is a prostate specific
XX peptide of the invention. The peptides either represent antigenic
XX epitopes or domains of prostate specific proteins.
XX
XX SQ Sequence 30 AA:

Query Match 100.0%; Score 148; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TAASDNFQLSGGGGFAIPGQAMAIAGQI 30
|||||
Db 1 taasdnfqlsggggfaipgqamaiaagqi 30

RESULT 2
AAU01143
ID AAU01143 standard; Protein: 30 AA.
XX

```

AC  AAM01143;
XX
XX  04-OCT-2001 (first entry)
XX
DE  Mycobacterium tuberculosis antigen Ra12 first 30 amino acids.
XX
XX  Human; prostate cancer; prostate-specific; diagnosis; vaccine;
XX  cytostatic; gene therapy; metastasis.
XX  Mycobacterium tuberculosis.
XX  WO200151633-A2.
XX  19-JUL-2001.
XX  16-JAN-2001; 2001WO-US01574.
XX  14-JAN-2000; 2000US-0483672.
XX
XX  (CORI-) CORIXA CORP.
XX  Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
XX  Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;
XX  Wang A, Meagher MJ;
XX  WPI; 2001-425873/45.
XX
XX  New polynucleotide encoding a prostate-specific protein, for
XX  diagnosing, monitoring and treating prostate cancer in a patient and
XX  for use in vaccines -
XX
XX  Example 17; Page 394; 543pp; English.
XX
XX  The present invention describes polynucleotide sequences (I) which encode
XX  prostate-specific proteins (II). (I) and (II) have cytostatic activity,
XX  and can be used in vaccine production and gene therapy. (I), (II),
XX  antibodies to (II), fusion proteins comprising (II), and isolated
XX  T cells prepared using (I) or (II) are used treat cancer in a
XX  patient. The cancer that is diagnosed or treated is particularly
XX  prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
XX  (I) can be used for monitoring the progression of cancer in a patient.
XX  (I) and (II) can also be used to improve diagnostic and therapeutic
XX  methods for prostate cancer. They can indicate the level of metastasis
XX  as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
XX  AAM01318 represent polynucleotide and amino acid sequences used in the
XX  exemplification of the present invention.
XX
XX  Sequence 30 AA:
XX
Query Match 100.0%; Score 148; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAASDNFQLSGGGFAIPICGAMAIAIGOI 30
    |||
Db 1 taasdnfqlsgggfaipigamalaigqi 30

```

```

RESULT 3
AAG99028
ID AAG99028 standard; Protein; 30 AA.
XX
AC AAG99028;
XX
XX 25-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis antigen Ra12 1st 30 amino acid sequence.
XX
XX Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
XX chromosome 22q11.2; prostate-specific protein; chromosome 1;
XX prostate specific antigen; PSA.
XX

```

```

XX
XX  Mycobacterium tuberculosis.
XX
XX  WO200134802-A2.
XX
XX  17-MAY-2001.
XX
XX  09-NOV-2000; 2000WO-US30904.
XX
XX  12-NOV-1999; 99US-0439313.
XX  18-NOV-1999; 99US-0443886.
XX
XX  (CORI-) CORIXA CORP.
XX
XX  Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
XX  Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAM, Wang A;
XX  WPI; 2001-308785/32.
XX
XX  Isolated polypeptide comprising at least an immunogenic portion of a
XX  prostate-specific protein, useful in the diagnosis and therapy of
XX  prostate cancer -
XX
XX  Example 17; Page 294; 325pp; English.
XX
XX  The present invention describes an isolated polypeptide (P1) comprising
XX  at least an immunogenic portion of a prostate-specific protein, or its
XX  variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
XX  (N1) have cytostatic activity and can be used in vaccine production.
XX  The polypeptides, nucleic acids and antibodies from the present
XX  invention are useful in the diagnosis and therapy of prostate cancer.
XX  Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
XX  in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
XX  region. Prostate specific antigen (PSA) P501S was located on
XX  chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
XX  polynucleotide and polypeptide sequences used in the exemplification
XX  of the present invention.
XX
XX  Sequence 30 AA:
XX
Query Match 100.0%; Score 148; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAASDNFQLSGGGFAIPICGAMAIAIGOI 30
    |||
Db 1 taasdnfqlsgggfaipigamalaigqi 30

```

Search completed: August 13, 2002, 10:59:49
Job time: 77 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:32 ; Search time 14.11 seconds
(without alignments)
204.300 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148

Sequence: 1 TAASDNFQLSQGGGQFAIPIGQAMAIAGQI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30

Maximum DB seq length: 30

Post-processing: Minimum Match 100%

Maximum Match 100%

Listing first 100 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match length	ID	Description
---------------	-----------------------	----	-------------

No matches found

Search completed: August 13, 2002, 10:59:14
Job time: 42 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:59:17 ; Search time 10.19 Seconds
(without alignments)
113.993 Million cell updates/sec

Title: US-09-684-215a-17
Perfect score: 148
Sequence: 1 TAASDNFQLSQGGGFAIPGQAMAIAQGI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: August 13, 2002, 11:02:53
Job time: 216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:58:57 ; Search time 23.81 seconds
(without alignments)
217.969 Million cell updates/sec

Title: US-09-684-215A-17

Perfect score: 148
Sequence: 1 TAASDNFQLSGCGCFARPICQAMAIAGOI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 30
Maximum DB seq length: 30

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 100 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_minc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
---------------	-------	----------------	----	-------------

No matches found

Search completed: August 13, 2002, 11:02:36
Job time: 219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:11 ; Search time 20.7 Seconds
(without alignments)
35.399 Million cell updates/sec

Title: US-09-684-215A-17
Perfect score: 148
Sequence: 1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	30	4 US-09-439-313-484	Sequence 484, Appl
2	148	100.0	132	4 US-08-818-112-66	Sequence 66, Appl
3	148	100.0	132	4 US-08-818-111-67	Sequence 67, Appl
4	148	100.0	132	4 US-09-056-556-66	Sequence 66, Appl
5	148	100.0	355	4 US-08-818-112-79	Sequence 79, Appl
6	148	100.0	355	4 US-08-818-111-80	Sequence 80, Appl
7	148	100.0	355	4 US-09-056-556-79	Sequence 79, Appl
8	55	37.2	580	4 US-08-818-112-75	Sequence 75, Appl
9	55	37.2	580	4 US-08-818-111-76	Sequence 76, Appl
10	55	37.2	580	4 US-09-056-556-75	Sequence 75, Appl
11	47.5	32.1	694	2 US-08-895-522-3	Sequence 3, Appl
12	47.5	32.1	694	3 US-09-195-391-3	Sequence 3, Appl
13	46.5	31.4	747	2 US-08-895-522-1	Sequence 1, Appl
14	46.5	31.4	747	3 US-09-195-391-1	Sequence 1, Appl
15	46	31.1	344	2 US-08-846-762-90	Sequence 90, Appl
16	44	29.7	460	4 US-09-199-637A-132	Sequence 132, App
17	43.5	29.4	776	1 US-07-603-133B-18	Sequence 18, Appl
18	43	29.1	1528	1 US-08-463-092B-6	Sequence 6, Appl
19	43	29.1	1528	2 US-08-463-092B-6	Sequence 6, Appl
20	43	29.1	1528	3 US-08-460-907B-6	Sequence 6, Appl
21	43	29.1	1528	2 US-08-463-179A-6	Sequence 6, Appl
22	43	29.1	1528	3 US-08-461-384B-6	Sequence 6, Appl
23	42.5	28.7	248	4 US-09-091-219-3	Sequence 3, Appl
24	42.5	28.7	2232	4 US-09-091-219-25	Sequence 25, Appl
25	42.5	28.7	2247	2 US-09-091-219-2	Sequence 2, Appl
26	42	28.4	323	2 US-08-874-347-21	Sequence 21, Appl
27	42	28.4	323	3 US-09-093-522-21	Sequence 21, Appl

ALIGNMENTS

28	42	28.4	547	1	US-08-340-203A-3	Sequence 3, Appl
29	42	28.4	547	2	US-08-452-567-3	Sequence 3, Appl
30	42	28.4	547	2	US-08-452-427-3	Sequence 3, Appl
31	42	28.4	547	3	US-09-085-407-3	Sequence 3, Appl
32	42	28.4	1243	2	US-08-557-139-2	Sequence 2, Appl
33	42	28.4	1334	2	US-08-996-545-2	Sequence 2, Appl
34	42	28.4	1334	4	US-09-328-320-2	Sequence 2, Appl
35	42	28.4	1531	1	US-08-141-893-2	Sequence 2, Appl
36	42	28.4	1531	1	US-08-463-092B-2	Sequence 2, Appl
37	42	28.4	1531	1	US-08-463-092B-2	Sequence 2, Appl
38	42	28.4	1531	2	US-08-462-109A-2	Sequence 2, Appl
39	42	28.4	1531	2	US-08-462-109A-2	Sequence 2, Appl
40	42	28.4	1531	2	US-08-460-907B-2	Sequence 2, Appl
41	42	28.4	1531	2	US-08-460-907B-4	Sequence 2, Appl
42	42	28.4	1531	3	US-08-463-179A-2	Sequence 2, Appl
43	42	28.4	1531	3	US-08-463-179A-2	Sequence 2, Appl
44	42	28.4	1531	3	US-08-461-384B-2	Sequence 2, Appl
45	42	28.4	1531	3	US-08-461-384B-4	Sequence 4, Appl

```
RESULT 1
US-09-439-313-484
; Sequence 484, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqul
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439.313
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-439-313-484

Query Match          100.0%; Score 148; DB 4; Length 30;
Best local similarity 100.0%; Pred. No. 1e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30
Db 1 TAASDNFOLSGGCGFAIPICGAMAIAGOI 30

RESULT 2
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
```

```

APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-66

Query Match          100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPIGQAMAIAGOI 30
Db 1 TAASDNFOLSGGGGFAIPIGQAMAIAGOI 30

RESULT 3
US-08-818-111-67
Sequence 67, Application US/08818111
Patent No. 633852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
```

```

FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-67

Query Match          100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPIGQAMAIAGOI 30
Db 1 TAASDNFOLSGGGGFAIPIGQAMAIAGOI 30

RESULT 4
US-09-056-556-66
Sequence 66, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-66

Query Match          100.0%; Score 148; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 6e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

TRE


```
QY 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30
    |||||
Db 1 TAASDNFQLSGGGGFAIPIGQAMAIAGQI 30
```

RESULT 5
US-08-818-112-79

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W
APPLICANT: Filler, David G

APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedick, Thomas S.
 APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: AND DIAGNOSIS
 ;
 ; NUMBER OF SEQUENCES: 153
 ;
 ; CORRESPONDENCE ADDRESS:
 ;
 ;

MEMORANDUM FOR THE DIRECTOR
SUBJECT: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
CONTACT: DR. JAMES R. HAYES, JR.

```

1  MEDIUM TYPE: Floppy disk
2
3  COMPUTER: IBM PC compatible
4
5  OPERATING SYSTEM: PC-DOS/MS-DOS
6
7  SOFTWARE: PatentIn Release #1.0,
8
9  CURRENT APPLICATION DATA:
10
11  APPLICATION NUMBER: US/08/818,112
12
13  Filing Date: 12-MAR-1987

```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
DECISIONATION NUMBER: 31 303

REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
FAX: (206) 622-6021

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:

US-08-818-112-75

Query Match	100.0%;	Score 148;	DB 4;	Length 355;
Best Local Similarly	100.0%;	Pred. No. 1.9e-14;		
Matches 30; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

```
QY      1 TAASDNFQLSOGGGAIPIGQAMAIAGOI 30  
        |||||  
Db     224 TAASDNFQLSOGGGAIPIGQAMAIAGOI 253
```

RESULT 6
US-08-818-111-80

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W

APPLICANT: Ellison, David C.
APPLICANT: Campos-Neto, Anton
APPLICANT: Houghton, Raymond
APPLICANT: Vedrick, Thomas S.
APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 148
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington

COMPUTER READABLE FORM:
MEDIUM TYPE: F]DDDD

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
RENEWAL DOCKET NUMBER: 310303 41766

TELECOMMUNICATION INFORMATION
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

US-08-818-111-80

Query Match	100.0%;	Score 148; DB 4;	Length 35;
Best Local Similarity	100.0%;	Pred. No. 1.9e-14;	
Matches 30; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0

```
QY      1 TAASDNFQLSQGGGFAIPIGAMAIAGQI   30  
         |||||  
Db     224 TAASDNFQLSQGGGFAIPIGAMAIAGQI   253
```

RESULT 7
US-09-056-556-79

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W

AFFILIANT: DILLON, DAVID C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 TREATMENT OF HIV INFECTION
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BENT LUT
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

```

COUNT:  CSR
;
; ZIP:  98104-7092
;
; COMPUTER READABLE FORM
;
; MEDIUM TYPE:  Floppy
;
; COMPUTER:  IBM PC co
;

```

```
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
;

```

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.45

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match          100.0%; Score 148; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNFOLSG--GCGFAIPGAMAIAGQI 30
Db 224 TAAADNFOLSGGCGFAIPGAMAIAGQI 253

RESULT 8
US-08-818-112-75
; Sequence 75, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-75

Query Match          37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. No. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAADNFOLSG--GCGFAIPGAMAIAGQI 30
```

```

Db 453 TLGADSDAQSGSIGLGFALPVDQAKRIADEL 484

RESULT 9
US-08-818-111-76
; Sequence 76, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-76

Query Match          37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. No. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAADNFOLSG--GCGFAIPGAMAIAGQI 30
Db 453 TLGADSDAQSGSIGLGFALPVDQAKRIADEL 484

RESULT 10
US-09-056-556-75
; Sequence 75, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
```

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-75

Query Match 37.2%; Score 55; DB 4; Length 580;
Best Local Similarity 40.6%; Pred. No. 2.5;
Matches 13; Conservative 5; Mismatches 12; Indels 2; Gaps 1;

QY 1 TAAASNFOLSOG--GQGFAIPGQAMAIAGQI 30
DB 453 TLGADSAQAQSSIGLGFALPVDQAKRIDEL 484

RESULT 11
US-08-895-522-3
Sequence 3, Application US/08895522
Patent No. 5858719
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
TITLE OF INVENTION: TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,522
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0336 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1167982
US-08-895-522-3

Query Match 32.1%; Score 47.5; DB 2; Length 694;
Best Local Similarity 37.1%; Pred. No. 41;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;

QY 1 TAAASN-FQLSOG-----GQGFAIPGQAMAIAG 28
DB 413 TVAFDNVHFYEYIEGQKVLNGVSFEVPAGKVAIVG 447

RESULT 12
US-09-195-391-3
Sequence 3, Application US/09195391
Patent No. 6080842
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
TITLE OF INVENTION: TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/195,391
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,522
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0336 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1167982
US-09-195-391-3

Query Match 32.1%; Score 47.5; DB 3; Length 694;
Best Local Similarity 37.1%; Pred. No. 41;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;
OY 1 TAASDN--FOLSG-----GQGAIPIGQAMA1G 28
DB 413 TVAFDNVHFYE1EGQKVLNGVSFEVPAKKAIVG 447

RESULT 13
US-08-895-522-1
; Sequence 1, Application US/08895522
; Patent No. 5858719
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
; TITLE OF INVENTION: TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,522
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0336 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: OVARNOT02
; CLONE: 545981
; US-08-895-522-1

Query Match 31.4%; Score 46.5; DB 2; Length 747;
Best Local Similarity 37.1%; Pred. No. 64;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;
OY 1 TAASDN--FOLSG-----GQGAIPIGQAMA1G 28
DB 466 TVAFDNVHFYE1EGQKVLNGVSFEVPAKKAIVG 500
RESULT 14
US-09-195-391-1
; Sequence 1, Application US/09195391
; Patent No. 6080842
; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ATP-BINDING CASSETTE
; TITLE OF INVENTION: TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,391
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/895,522
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0336 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: OVARNOT02
; CLONE: 545981
; US-09-195-391-1

Query Match 31.4%; Score 46.5; DB 3; Length 747;
Best Local Similarity 37.1%; Pred. No. 64;
Matches 13; Conservative 5; Mismatches 10; Indels 7; Gaps 2;
OY 1 TAASDN--FOLSG-----GQGAIPIGQAMA1G 28
DB 466 TVAFDNVHFYE1EGQKVLNGVSFEVPAKKAIVG 500

RESULT 15
US-08-846-762-90
; Sequence 90, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 90
; LENGTH: 344
; TYPE: PRT

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:35:41 ; Search time 20.7 Seconds
(without alignments)
151.037 Million cell updates/sec

Title: US-09-684-215A-18

Perfect score: 653
Sequence: 1 TAAADNFQLSGGOGFAIP1.....QTKSGTFTGNVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCPUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	627	96.0	132	4	US-08-818-112-66 Sequence 66, Appl
2	627	96.0	132	4	US-08-818-111-67 Sequence 67, Appl
3	627	96.0	132	4	US-09-056-556-66 Sequence 66, Appl
4	627	96.0	335	4	US-08-818-112-79 Sequence 79, Appl
5	627	96.0	335	4	US-08-818-111-80 Sequence 80, Appl
6	627	96.0	335	4	US-09-056-556-79 Sequence 79, Appl
7	201.5	30.9	580	4	US-08-818-112-75 Sequence 75, Appl
8	201.5	30.9	580	4	US-08-818-111-76 Sequence 76, Appl
9	201.5	30.9	580	4	US-09-056-556-75 Sequence 75, Appl
10	158.5	24.3	97	4	US-08-818-112-72 Sequence 72, Appl
11	158.5	24.3	97	4	US-08-818-111-73 Sequence 73, Appl
12	158.5	24.3	97	4	US-09-056-556-72 Sequence 72, Appl
13	148	22.7	30	4	US-09-439-313-484 Sequence 484, App
14	128.5	19.7	460	4	US-09-199-637A-132 Sequence 132, App
15	106	16.2	475	1	US-08-278-091-6 Sequence 6, Appli
16	106	16.2	475	1	US-08-483-859-6 Sequence 6, Appli
17	106	16.2	475	1	US-08-472-173-6 Sequence 6, Appli
18	106	16.2	475	1	US-08-350-741-2 Sequence 2, Appli
19	106	16.2	475	1	US-08-487-167-6 Sequence 6, Appli
20	106	16.2	475	2	US-08-482-816-6 Sequence 6, Appli
21	106	16.2	475	2	US-08-296-149-6 Sequence 6, Appli
22	106	16.2	475	2	US-08-801-499-6 Sequence 6, Appli
23	106	16.2	475	2	US-08-463-875A-2 Sequence 6, Appli
24	106	16.2	475	2	US-08-615-271-6 Sequence 6, Appli
25	106	16.2	475	3	US-09-074-660-6 Sequence 6, Appli
26	106	16.2	475	3	US-09-074-659-6 Sequence 6, Appli
27	106	16.2	475	3	US-09-106-468-6 Sequence 6, Appli

28	106	16.2	475	4	US-09-106-466A-6 Sequence 6, Appli
29	106	16.2	475	4	US-09-106-467-6 Sequence 6, Appli
30	103.5	15.8	472	1	US-08-278-091-5 Sequence 5, Appli
31	103.5	15.8	472	1	US-08-483-859-5 Sequence 5, Appli
32	103.5	15.8	472	1	US-08-472-173-5 Sequence 5, Appli
33	103.5	15.8	472	2	US-08-487-167-5 Sequence 5, Appli
34	103.5	15.8	472	2	US-08-482-816-5 Sequence 5, Appli
35	103.5	15.8	472	2	US-08-296-149-5 Sequence 5, Appli
36	103.5	15.8	472	2	US-08-801-499-5 Sequence 5, Appli
37	103.5	15.8	472	2	US-08-615-271-5 Sequence 5, Appli
38	103.5	15.8	472	3	US-09-074-660-5 Sequence 5, Appli
39	103.5	15.8	472	3	US-09-074-659-5 Sequence 5, Appli
40	103.5	15.8	472	3	US-09-106-468-5 Sequence 5, Appli
41	103.5	15.8	472	4	US-09-106-466A-5 Sequence 5, Appli
42	103.5	15.8	472	4	US-09-106-467-5 Sequence 5, Appli
43	103.5	15.8	491	3	US-08-923-454A-14 Sequence 14, Appli
44	103.5	15.8	492	1	US-08-350-741-3 Sequence 3, Appli
45	103.5	15.8	492	2	US-08-463-875A-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-818-112-66
Sequence 66, Application US/08818112
Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

US-08-818-112-66

Query Match 96.0%; Score 627; DB 4; Length 132;
Best Local Similarity 94.7%; Pred. No. 2, 4e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY	1	TAASDNPDLGGGGGFAIPICQAMALIAQK---- <td>56</td>	56
Db	1	TAASDNPDLGGGGGFAIPICQAMALIAQKSGGSPETHIGPTAFLGLGVYDNNNGAR	60
QY	57	VORVVGSAFASLSLGTSDVYTTAVDGAIPNSATMADALNGHPGDVLSYTWQTKSGGTR	116
Db	61	VORVVGSAFASLSLGTSDVYTTAVDGAIPNSATMADALNGHPGDVLSYTWQTKSGGTR	120
QY	117	TGNVTIAEGPPA	128
Db	121	TGNVTIAEGPPA	132

```

1      RESULT 2
2      US-08-818-111-67
3      ; Sequence 67, Application US/06818111
4      ; Patent No. 633852
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Reed, Steven G.
7      ; APPLICANT: Skelky, Yasir A.W.
8      ; APPLICANT: Dillon, David C.
9      ; APPLICANT: Campos-Neto, Antonia
10     ; APPLICANT: Houghton, Raymond
11     ; APPLICANT: Vedavick, Thomas S.
12     ; APPLICANT: Wardzik, Daniel R.
13     ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
14     ; NUMBER OF SEQUENCES: 148
15     ; CORRESPONDENCE ADDRESS:
16     ; ADDRESSEE: SEED and BERRY LLP
17     ; STREET: 6300 Columbia Center, 701 Fifth Avenue
18     ; CITY: Seattle
19     ; STATE: Washington
20     ; COUNTRY: USA
21     ; ZIP: 98104-7092
22     ; COMPUTER READABLE FORM:
23     ; MEDIUM TYPE: Floppy disk
24     ; COMPUTER: IBM PC compatible
25     ; OPERATING SYSTEM: PC-DOS/MS-DOS
26     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
27     ; CURRENT APPLICATION DATA:
28     ; APPLICATION NUMBER: US/08/818,111
29     ; FILING DATE: 13-MAR-1997
30     ; CLASSIFICATION: 424
31     ; ATTORNEY/AGENT INFORMATION:
32     ; NAME: Makl, David J.
33     ; REGISTRATION NUMBER: 31,392
34     ; REFERENCE/DOCKET NUMBER: 210121.417C6
35     ; TELECOMMUNICATION INFORMATION:
36     ; TELEPHONE: (206) 622-4900
37     ; TELEFAX: (206) 682-6031
38     ; INFORMATION FOR SEQ ID NO: 67:
39     ; SEQUENCE CHARACTERISTICS:
40     ; LENGTH: 132 amino acids
41     ; TYPE: amino acid
42     ; STRANDEDNESS: single
43     ; TOPOLOGY: linear
44     ;
45     ;
46     ;
47     ;
48     ;
49     ;
50     ;
51     ;
52     ;
53     ;
54     ;
55     ;
56     ;
57     ;
58     ;
59     ;
60     ;
61     ;
62     ;
63     ;
64     ;
65     ;
66     ;
67     ;
68     ;
69     ;
70     ;
71     ;
72     ;
73     ;
74     ;
75     ;
76     ;
77     ;
78     ;
79     ;
80     ;
81     ;
82     ;
83     ;
84     ;
85     ;
86     ;
87     ;
88     ;
89     ;
90     ;
91     ;
92     ;
93     ;
94     ;
95     ;
96     ;
97     ;
98     ;
99     ;
100    ;
101    ;
102    ;
103    ;
104    ;
105    ;
106    ;
107    ;
108    ;
109    ;
110    ;
111    ;
112    ;
113    ;
114    ;
115    ;
116    ;
117    ;
118    ;
119    ;
120    ;
121    ;
122    ;
123    ;
124    ;
125    ;
126    ;
127    ;
128    ;
129    ;
130    ;
131    ;
132    ;
133    ;
134    ;
135    ;
136    ;
137    ;
138    ;
139    ;
140    ;
141    ;
142    ;
143    ;
144    ;
145    ;
146    ;
147    ;
148    ;
149    ;
150    ;
151    ;
152    ;
153    ;
154    ;
155    ;
156    ;
157    ;
158    ;
159    ;
160    ;
161    ;
162    ;
163    ;
164    ;
165    ;
166    ;
167    ;
168    ;
169    ;
170    ;
171    ;
172    ;
173    ;
174    ;
175    ;
176    ;
177    ;
178    ;
179    ;
180    ;
181    ;
182    ;
183    ;
184    ;
185    ;
186    ;
187    ;
188    ;
189    ;
190    ;
191    ;
192    ;
193    ;
194    ;
195    ;
196    ;
197    ;
198    ;
199    ;
200    ;
201    ;
202    ;
203    ;
204    ;
205    ;
206    ;
207    ;
208    ;
209    ;
210    ;
211    ;
212    ;
213    ;
214    ;
215    ;
216    ;
217    ;
218    ;
219    ;
220    ;
221    ;
222    ;
223    ;
224    ;
225    ;
226    ;
227    ;
228    ;
229    ;
230    ;
231    ;
232    ;
233    ;
234    ;
235    ;
236    ;
237    ;
238    ;
239    ;
240    ;
241    ;
242    ;
243    ;
244    ;
245    ;
246    ;
247    ;
248    ;
249    ;
250    ;
251    ;
252    ;
253    ;
254    ;
255    ;
256    ;
257    ;
258    ;
259    ;
260    ;
261    ;
262    ;
263    ;
264    ;
265    ;
266    ;
267    ;
268    ;
269    ;
270    ;
271    ;
272    ;
273    ;
274    ;
275    ;
276    ;
277    ;
278    ;
279    ;
280    ;
281    ;
282    ;
283    ;
284    ;
285    ;
286    ;
287    ;
288    ;
289    ;
290    ;
291    ;
292    ;
293    ;
294    ;
295    ;
296    ;
297    ;
298    ;
299    ;
300    ;
301    ;
302    ;
303    ;
304    ;
305    ;
306    ;
307    ;
308    ;
309    ;
310    ;
311    ;
312    ;
313    ;
314    ;
315    ;
316    ;
317    ;
318    ;
319    ;
320    ;
321    ;
322    ;
323    ;
324    ;
325    ;
326    ;
327    ;
328    ;
329    ;
330    ;
331    ;
332    ;
333    ;
334    ;
335    ;
336    ;
337    ;
338    ;
339    ;
340    ;
341    ;
342    ;
343    ;
344    ;
345    ;
346    ;
347    ;
348    ;
349    ;
350    ;
351    ;
352    ;
353    ;
354    ;
355    ;
356    ;
357    ;
358    ;
359    ;
360    ;
361    ;
362    ;
363    ;
364    ;
365    ;
366    ;
367    ;
368    ;
369    ;
370    ;
371    ;
372    ;
373    ;
374    ;
375    ;
376    ;
377    ;
378    ;
379    ;
380    ;
381    ;
382    ;
383    ;
384    ;
385    ;
386    ;
387    ;
388    ;
389    ;
390    ;
391    ;
392    ;
393    ;
394    ;
395    ;
396    ;
397    ;
398    ;
399    ;
400    ;
401    ;
402    ;
403    ;
404    ;
405    ;
406    ;
407    ;
408    ;
409    ;
410    ;
411    ;
412    ;
413    ;
414    ;
415    ;
416    ;
417    ;
418    ;
419    ;
420    ;
421    ;
422    ;
423    ;
424    ;
425    ;
426    ;
427    ;
428    ;
429    ;
430    ;
431    ;
432    ;
433    ;
434    ;
435    ;
436    ;
437    ;
438    ;
439    ;
440    ;
441    ;
442    ;
443    ;
444    ;
445    ;
446    ;
447    ;
448    ;
449    ;
450    ;
451    ;
452    ;
453    ;
454    ;
455    ;
456    ;
457    ;
458    ;
459    ;
460    ;
461    ;
462    ;
463    ;
464    ;
465    ;
466    ;
467    ;
468    ;
469    ;
470    ;
471    ;
472    ;
473    ;
474    ;
475    ;
476    ;
477    ;
478    ;
479    ;
480    ;
481    ;
482    ;
483    ;
484    ;
485    ;
486    ;
487    ;
488    ;
489    ;
490    ;
491    ;
492    ;
493    ;
494    ;
495    ;
496    ;
497    ;
498    ;
499    ;
500    ;
501    ;
502    ;
503    ;
504    ;
505    ;
506    ;
507    ;
508    ;
509    ;
510    ;
511    ;
512    ;
513    ;
514    ;
515    ;
516    ;
517    ;
518    ;
519    ;
520    ;
521    ;
522    ;
523    ;
524    ;
525    ;
526    ;
527    ;
528    ;
529    ;
530    ;
531    ;
532    ;
533    ;
534    ;
535    ;
536    ;
537    ;
538    ;
539    ;
540    ;
541    ;
542    ;
543    ;
544    ;
545    ;
546    ;
547    ;
548    ;
549    ;
550    ;
551    ;
552    ;
553    ;
554    ;
555    ;
556    ;
557    ;
558    ;
559    ;
560    ;
561    ;
562    ;
563    ;
564    ;
565    ;
566    ;
567    ;
568    ;
569    ;
570    ;
571    ;
572    ;
573    ;
574    ;
575    ;
576    ;
577    ;
578    ;
579    ;
580    ;
581    ;
582    ;
583    ;
584    ;
585    ;
586    ;
587    ;
588    ;
589    ;
590    ;
591    ;
592    ;
593    ;
594    ;
595    ;
596    ;
597    ;
598    ;
599    ;
600    ;
601    ;
602    ;
603    ;
604    ;
605    ;
606    ;
607    ;
608    ;
609    ;
610    ;
611    ;
612    ;
613    ;
614    ;
615    ;
616    ;
617    ;
618    ;
619    ;
620    ;
621    ;
622    ;
623    ;
624    ;
625    ;
626    ;
627    ;
628    ;
629    ;
630    ;
631    ;
632    ;
6
```

Query Match	Best Local Similarity	Score 627;	DB 4;	Length 132;
Matchs 125;	Conservative 1;	Pred. No. 2,4e-65;	Mismatches 2;	Indels 4;
				Gaps 1;
QY	1	TAASNFOLSOOGGOFALPIPGOMIAAGIK----	LPHYHIGPTAFLGIGVDDNNGNGAR	56
Db	1	TAASNFOLSOOGGOFALPIPGOMIAAQIRSGGSPVTHIGPTAFLGIGVDDNNGNGAR		60
QY	57	VQRVVGSAPASLIGISTGDVITAVYVGAPINSTATMAADLNGHHPPDVI	SVTWQTSKGR	116
Db	61	VQRVVGSAPASLIGISTGDVITAVYVGAPINSTATMAADLNGHHPPDVI	SVTWQTSKGR	120
QY	117	TGNVTLAEGPPA		128

```

Db          121 TGNVTLAEGBPA 132

|||||
|||
RESULT      3
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasar A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-Apr-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

```

```

Query Match          96.0%; Score 627; DB 4; Length 132;
Best Local Similarity 94.7%; Pred. No. 2,4e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1.

QY      1  TAASNPFLDSOGGSGFAFLPIGOMAMIAAGIK---LPVHIGPFLAFLIGGVYDNNNGNGAR 56
Db      1  TAASNPFLDSOGGSGFAFLPIGOMAMIAAGIRSGGSPVYHISLPFLAFLIGGVYDNNNGNGAR 60
QY      57  VQRVVGSAPAAASLGISTDVITAVDGAFLNSATAMADALNGHHPPSDVITSVTQTKSGGTR 116
Db      61  VQRVVGSAPAAASLGISTDVITAVDGAFLNSATAMADALNGHHPPSDVITSVTQTKSGGTR 120
QY      117 TGNVTLLAEGPPA 128
Db      121 TGNVTLLAEGPPA 132

RESULT      4
US-08-818-112-79
Sequence 79, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond

```



```

; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match          96.0%; Score 627; DB 4; Length 355;
Best Local Similarity 94.7%; Pred. No. 9,6e-65;
Matches 125; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 TASDNFQLSOGGQGFAPIGQAMATIGQK----LPTVHIGPPIAFGLGVYDNNNGNGAR 56
DB 224 TASDNFQLSOGGQGFAPIGQAMATIGQKRSGGSPVHIGPPIAFGLGVYDNNNGNGAR 283

QY 57 VORVVGSAAPASLIGSTGDTAVDGAIPNSATAMADALNGHHPGDVISTVWOTKSGSTR 116
DB 284 VORVVGSAAPASLIGSTGDTAVDGAIPNSATAMADALNGHHPGDVISTVWOTKSGSTR 343

QY 117 TGNVTTLAEGRPA 128
DB 344 TGNVTTLAEGRPA 355

RESULT 7
US-08-818-112-75
; Sequence 75, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-76
```

```

; TOPOLOGY: linear
US-08-818-112-75

Query Match          30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4,2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TASDNFQLSOG--GQGFAPIGQAMATIGQKLPVHIGPPIAFGLGVY---DNNNGNGA 55
DB 453 TLGADSDAQAQSGSIGGFAIPVDQAKRIADEL---ISTGKASHASIGVQVTDNDKTPGA 508

QY 56 RYQRYVGSAPASLIGSTGDTAVDGAIPNSATAMADALNGHHPGDVISTVWOTKSGSTR 115
DB 509 KIVEYVAGGAANAGVPKGVVYTKDDRPINSADALVAARSKAPATVALTFQDPDSGGS 568

QY 116 RTGVNVTTLAE 124
DB 569 RTGVNVTTLAE 577

RESULT 8
US-08-818-111-76
; Sequence 76, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-76

Query Match          30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4,2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TASDNFQLSOG--GQGFAPIGQAMATIGQKLPVHIGPPIAFGLGVY---DNNNGNGA 55
DB 453 TLGADSDAQAQSGSIGGFAIPVDQAKRIADEL---ISTGKASHASIGVQVTDNDKTPGA 508
```

```

Db 453 TLGADSAQSGSIGLGFAPVQAKRIADEL-----ISTGKASHASLGQVYTNDRKDPGA 508
OY 56 RVORVGSAPASLIGSDVITAVDAPINSATAMALNGHGGDVISYTWQKSGT 115
Db 509 KIVEVAGGAANAGVPGVYTKVDRPINSADALVAAVRSKAPGATVALTFQDPGSGS 568
OY 116 RTGNVTILAE 124
Db 569 RTVOVTLGK 577

RESULT 9
US-09-056-556-75
; Sequence 75, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-75

Query Match 30.9%; Score 201.5; DB 4; Length 580;
Best Local Similarity 38.0%; Pred. No. 4,2e-15;
Matches 49; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

OY 1 TAAADNFOLSG--GGFAIPIGQAMAIAGQIKLPVHIGPTAFLGLGV--DNNNGNGA 55
Db 453 TLGADSAQSGSIGLGFAPVQAKRIADEL-----ISTGKASHASLGQVYTNDRKDPGA 508
OY 56 RVORVGSAPASLIGSDVITAVDAPINSATAMALNGHGGDVISYTWQKSGT 115
Db 509 KIVEVAGGAANAGVPGVYTKVDRPINSADALVAAVRSKAPGATVALTFQDPGSGS 568
OY 116 RTGNVTILAE 124
Db 569 RTVOVTLGK 577

RESULT 10
US-08-818-112-72
; Sequence 72, Application US/08818112

```

```

; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-72

Query Match 24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3,3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

OY 35 VHTGPTAFGLGV--DNNNGCARVORVVGSAAPASLIGSDVITAVDAPINSATAM 91
Db 2 ISTGKASHASLGQVYTNDRKDPGAKIYEYVAGGAANAGVPGVYTKVDRPINSADAL 61
OY 92 ADALNGHHPGDVISTYTWQKSGTGTGNVTILAE 124
Db 62 VAAVRSKAPGATVALTFQDPGSGSRVQVTLGK 94

RESULT 11
US-08-818-111-73
; Sequence 73, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; TUBERCULOSIS

```

```

; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-818-111-73

Query Match          24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3.3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

QY 35 VHIGPTAFGLGV--DNGNGARVQRYGSA PASLIGSTGDTTAVDGA PINSATAM 91
; | : | | | | | : | | : | | | | | : | : | | | | | | | | :
Db 2 ISTGASHASLGVQYTNKDPGAKIVEYVAGGAANAGVPKGVVTKVDPRINSADAL 61

QY 92 ADALNGHHPGDVISTWQTKSGTGTGNVTLAE 124
; | : | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 VAAVRSKAPGATVALTFO DPSGSGSRVQVTLGK 94

RESULT 12
US-09-056-556-72
; Sequence 72, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-056-556-72

Query Match          24.3%; Score 158.5; DB 4; Length 97;
Best Local Similarity 38.7%; Pred. No. 3.3e-11;
Matches 36; Conservative 16; Mismatches 38; Indels 3; Gaps 1;

QY 35 VHIGPTAFGLGV--DNGNGARVQRYGSA PASLIGSTGDTTAVDGA PINSATAM 91
; | : | | | | | : | | : | | | | | : | : | | | | | | | | :
Db 2 ISTGASHASLGVQYTNKDPGAKIVEYVAGGAANAGVPKGVVTKVDPRINSADAL 61

QY 92 ADALNGHHPGDVISTWQTKSGTGTGNVTLAE 124
; | : | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 VAAVRSKAPGATVALTFO DPSGSGSRVQVTLGK 94

RESULT 13
US-09-439-313-484
; Sequence 484, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqun
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ. ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ. ID NO 484
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Sapien
;
US-09-439-313-484

Query Match          22.7%; Score 148; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFQLSGGGFAIPIGQMAAINGOI 30
; | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 TAASDNFQLSGGGFAIPIGQMAAINGOI 30

RESULT 14
US-09-199-637A-132
; Sequence 132, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:36:14 ; Search time 25.01 Seconds
(without alignments)
491.781 Million cell updates/sec

Title: US-09-684-215a-18

Perfect score: 653
Sequence: 1 TAAADNFOLSGGGGFAIRP.....QTKSGTRNGVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR71:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632	96.8	355	2 F70983	probable serine pr
2	437.5	67.0	361	2 S47170	hypothetical prote
3	422.5	64.7	354	2 A87242	probable secreted
4	200.5	30.7	464	2 C70821	probable serine pr
5	185.5	28.4	382	2 H86930	probable secreted
6	185.5	28.4	452	2 T45348	probable serine pr
7	140	21.4	407	2 AG2150	serine proteinase
8	134.5	20.6	362	2 T35287	exported serine pr
9	132.5	20.3	408	2 H86891	serine proteinase
10	123.5	18.9	474	2 F83550	periplasmic protei
11	122.5	18.7	514	2 A82851	proteinase hha (E
12	122	18.7	394	2 S74643	proteinase hha (E
13	119	18.2	441	2 E75357	probable periplasm
14	116.5	17.8	203	2 T35866	probable integral
15	115	17.6	348	2 H96956	serine protease Do
16	112.5	17.2	530	2 F87590	serine proteinase
17	111	17.0	499	2 B81914	probable periplasm
18	110	16.8	455	2 C91142	serine endoprotein
19	110	16.8	455	2 F85987	serine endoprotein
20	109	16.7	393	2 E95261	serine proteinase
21	109	16.7	397	2 B98127	serine proteinase
22	108	16.5	429	2 AD1894	serine proteinase
23	107	16.4	513	2 I40060	serine proteinase
24	107	16.4	513	2 AD3418	proteinase DO (EC
25	106.5	16.3	459	2 F72359	periplasmic serine
26	106	16.2	455	2 AB0909	serine protease (E
27	106	16.2	475	1 S15317	heat shock protein
28	105.5	16.2	355	1 JC6052	trypsin-like prote
29	105.5	16.2	355	2 D91142	proteinase (import

30	105.5	16.2	355	2 G85987	proteinase (import
31	105.5	16.2	452	2 S77538	serine proteinase
32	105	16.1	455	2 JC6051	trypsin-like prote
33	104	15.9	475	2 AC0528	proteinase DO precu
34	103.5	15.8	474	2 S45229	proteinase DO (EC
35	103.5	15.8	474	2 B85500	proteinase DO (EC
36	103.5	15.8	474	2 E90649	proteinase DO (EC
37	103	15.8	523	2 A97479	probable serine pr
38	103	15.8	523	2 A12696	serine proteinase
39	102.5	15.7	404	2 C70886	probable integral
40	102	15.6	472	2 C87408	serine proteinase
41	102	15.6	476	2 H71936	proteinase DO - He
42	101	15.5	398	2 B71284	proteinase DO (EC
43	98.5	15.1	362	2 A10433	proteinase (EC 3.4
44	98.5	15.1	481	2 D82826	heat shock protein
45	97.5	14.9	356	2 AC0909	serine protease (E

ALIGNMENTS

RESULT 1
F70983
probable serine proteinase pepa - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R: Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID: 98295987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-355 <COL>
A: Cross-references: GB:296071; GB:AL123456; NID:93242254; PIDN: CAB09453.1; PID: g21819
A: Experimental source: Strain H37RV
C: Genetics:
A: Gene: pepa
C: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 96.8%; Score 632; DB 2; Length 355;
Best Local Similarity 95.3%; Pred. No. 1.3e-47;
Matches 126; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

OY 1 TAAADNFOLSGGGGFAIRP...GAMAIGQIK---LPTVHIGPTAFGLGVVNNNGNGAR 56
|||||
Db 224 TAAADNFOLSGGGGFAIRP...GAMAIGQIRSGGSPVHIGPTAFGLGVVNNNGNGAR 283
OY 57 VQRYVGSAPASLIGSTGDIYTAVDGAPINSATAMADALNCHHFGDYSVTWQKSGSTR 116
|||||
Db 284 VQRYVGSAPASLIGSTGDIYTAVDGAPINSATAMADALNCHHFGDYSVTWQKSGSTR 343
OY 117 TGNVTLAEGPPA 128
|||||
Db 344 TGNVTLAEGPPA 355

RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
Submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34Kda protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170

Db 381 RDGKSNFADVKLSK 394

RESULT 10

F83550
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83550
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337
 A:Accession: F83550
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1474 <STD>
 A:Cross-references: GB:AE004511; GB:AE004091; NID:99946646; PIDN:AA04155.1; GSPDB:GN001
 C:Genetics:
 A:Experimental source: strain PA01
 A:Gene: mucD; PA0766

Query Match 18.9%; Score 123.5; DB 2; Length 474;
 Best Local Similarity 31.6%; Pred. No. 0.0027;
 Matches 43; Conservative 17; Mismatches 53; Indels 23; Gaps 4;

QY 4 SDNPLSOG--GQGFAPRIGQAMAIAGQIKLPTVHIGPTAFGLGVVDNNGN-----53

Db 233 SQIFRSGGFMLSPFAIPDVALNADQK---KAGKSRGWLIVQEVNKDLAESFG 288

QY 54 -----GAVQRYVGSAPASLIGSTGDTTAVDAPINGATAMADLNGHHPDVSF--106

Db 289 LDKPSGALVADLVEDGPPAKGSLQYGVYLSLNGSINSADLPHLVGNKPKEDKINLDV 348

QY 107 --TWQKSGGRTGNV 120

Db 349 IRNGKRSLSMAVGSL 364

RESULT 11

A82581
 periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82581
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-514 <STM>
 A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAF8040.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H. as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kurama, E.E.; Laig Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak M.; Tshako, M.R.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
 A:Contents: annotation

C:Genetics:
 A:Gene: XF2241
 C:Superfamily: Helicobacter serine proteinase

Query Match 18.8%; Score 122.5; DB 2; Length 514;
 Best Local Similarity 36.9%; Pred. No. 0.0036;
 Matches 41; Conservative 13; Mismatches 42; Indels 15; Gaps 4;

QY 4 SDNPLSOG--GQGFAPRIGQAMAIAGQIK-----LPTVHIGPTAF--GLGVVDN 50

Db 262 SQIFASGCGWGISFAIPDVALNADQK---KAGKSRGWLIVQEVNKDLAESFG 321

QY 51 NGNGARQRYVGSAPASLIGSTGDTTAVDAPINGATAMADLNGHHPG 101

Db 322 --RGALVNNIPHPSPAPAKGIEVGDVIRSVNGKVISFSFDLPPLIGMPPG 370

RESULT 12

S74643
 proteinase hbaA (EC 3.4.-.-) - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein sll1679
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis.
 A:Reference number: S74322; MUID:97061201

A:Accession: S74643
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-394 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:91651768; PIDN:BA16795.1; PID:9165
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: hbaA
 C:Superfamily: proteinase hbaB
 C:Keywords: hydrolase; proteinase

Query Match 18.7%; Score 122; DB 2; Length 394;
 Best Local Similarity 32.8%; Pred. No. 0.003;
 Matches 43; Conservative 11; Mismatches 39; Indels 38; Gaps 5;

QY 13 GQGFAPRIGQAMAIAGQIKLPTVHIGPTAFGLGVVD-----NNGN-----53

Db 261 GIGFAIPIDQAKAI--QNTLAAGTVPHPYIGVGMNMTVDQAQNNRNPSPFIPEVD 318

QY 54 GAVQRYVGSAPASLIGSTGDTTAVDAPINGATAMADLNGHHPG 99

Db 319 GILVNRVLPETPARAGIRRGDIYAVDGPISDGAIRQIRVEAGINKALKLIDLRLDR 378

QY 100 PGDVISVTWQT 110

Db 379 R---LSLTWQT 386

RESULT 13

E75357
 Probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75357
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J. M.; Shen, M.; Vanattervan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896

A:Accession: E75357
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <WHI>
A:Cross-references: GB:AE002017; GB:AE000513; NID:96459527; PIDN:AAFI1312.1; PID:9645953
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1756
A:Map position: 1
C:Superfamily: proteinase hnoB

Query Match 18.2%; Score 119; DB 2; Length 441;
Best Local Similarity 28.6%; Pred. No. 0.0061;
Matches 46; Conservative 22; Mismatches 43; Indels 50; Gaps 7;

QY 6 NQIOLSGGQ-----GFAIPGQAMAIAGQIKLPYHIGPTAFL---GLGYVDNN-----51
DB 277 NQIOLGAGAGAGAGVGFALPINTVKRL---LPQLQAGKGGVSPSPSGVPSDLSLP 331
QY 52 -----GNGARVQRYVGSAPASLIG-----IST-GDVITAVD 81
DB 332 PQLKRAAGLPSSGALLQKYPGSPALAAAGLRGNNKGLSPNQTSISTDGLITAVN 391
QY 82 GAPINSATAMADALNGHHPGDVISTVQTKSGGTRGVNTL 122
DB 392 GQPLEDAGSIQEAIVLATGEGQPLRTVR--RGKTRREVETL 431

RESULT 14

T35866
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T35866
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A:Reference number: 221591
A:Accession: T35866
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-203 <SAU>
A:Cross-references: EMBL:AL049727; PIDN:CA041567.1; GSPDB:GN00070; SCOEDB:SC9B1.21
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC9B1.21

Query Match 17.8%; Score 116.5; DB 2; Length 203;
Best Local Similarity 32.5%; Pred. No. 0.0043;
Matches 37; Conservative 15; Mismatches 49; Indels 13; Gaps 2;

QY 12 GGCGFAIPGQAMAIAGQIKLPYHIGPT-----AFLGLGVVDNNGNGARVQRY 60
DB 80 GGAGGCGATGTATGTRGAPAHPSA--GPSAGRSPSPACATLGEVAVDDEKPGARVGV 137
QY 61 VGSAPASLIGISTGDVITAVDAPINSATAMADALNGHHPGDVISTVQTKSGG 114
DB 138 HVPQPGYAAGLVNGDVLLAVGTRVDSATDLAHAVARAGPEKVKLTVRHRSRG 191

RESULT 15

H96936
serine protease Do (heat-shock protein) [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H96936
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H96936

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78443.1; PID:91502321; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0463

Query Match 17.6%; Score 115; DB 2; Length 348;
Best Local Similarity 28.2%; Pred. No. 0.011;
Matches 35; Conservative 25; Mismatches 48; Indels 16; Gaps 4;

QY 10 SQGGGFAIPG-----QAMAIAGQIKLPYHI-----GTFALGLGVVDNNGNGARVQ 58
DB 226 SAEIGFAVPIINIVKRYLSKTKTGQFKTPVIGIIGLDSMNGVLTN---NFEKGIYV 281
QY 59 RVVGSAPASLIGISTGDVITAVDAPINSATAMADALNGHHPGDVISTVQTKSGGTRTG 118
DB 282 NISPSGAAAGINKGDIILSVNGKNININNELRESIYTTGANNVSLKIKTAS-GEKTV 340
QY 119 NVTL 122
DB 341 NVKI 344

Search completed: August 13, 2002, 10:36:15
Job time: 63 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:42:45 ; Search time 15.81 Seconds
(without alignments)
313.479 Million cell updates/sec

Title: US-09-684-215a-18

Perfect score: 653
Sequence: 1 TPAASNPQLSGGCGFAIP1.....QTKSGTRTGNVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	20.3	408	1	HTRA_LACTA
2	129	19.8	413	1	HTRA_LACTA
3	114	17.5	504	1	DEGP_RHME
4	107	16.4	513	1	DEGP_BRUB
5	106	16.2	475	1	DEGP_SALT
6	105.5	16.2	355	1	DEGS_ECOLI
7	105	16.1	455	1	DEGQ_ECOLI
8	103.5	15.8	474	1	DEGP_ECOLI
9	103.5	15.8	503	1	DEGP_BARHE
10	94	14.4	478	1	DEGP_BUCAI
11	92	14.1	448	1	DEGP_ARANTH
12	88	13.5	497	1	DEGP_CHTR
13	88	13.5	905	1	ZOJ_MOUSE
14	87.5	13.4	466	1	HTOA_HABIN
15	86	13.2	510	1	MURE_MYCTU
16	85.5	13.1	492	1	NYLA_PSES8
17	85.5	13.1	492	1	NYLA_PSES8
18	85.5	13.1	492	1	DEGP_CHIMU
19	84	12.9	864	1	ELC_RAT
20	83	12.7	400	1	YTXA_BACSU
21	82	12.6	762	1	P115_CHICK
22	81	12.4	488	1	DEGP_CHIPN
23	79.5	12.2	292	1	PPNK_ECOLI
24	79.5	12.2	292	1	PPNK_ECOLI
25	79	12.1	2249	1	OMPA_RICRI
26	78.5	12.0	437	1	DEGL_ARANTH
27	78.5	12.0	543	1	7UP2_DROME
28	78.5	12.0	746	1	7UP2_DROME
29	77.5	11.9	460	1	HRA3_MOUSE
30	77.5	11.9	478	1	DEGP_BUCAP
31	77	11.8	430	1	TPSN_CHICK
32	77	11.8	2021	1	OMPA_RICRI
33	76.5	11.7	219	1	VG36_BPT2

34	76.5	11.7	928	1	PMF9_CHLPN
35	75.5	11.6	475	1	TRB1_ECOLI
36	75.5	11.6	852	1	DLG2_RAT
37	75.5	11.6	1286	1	AIDA_ECOLI
38	75	11.5	467	1	ISP6_SCHPO
39	74.5	11.4	351	1	HYPE_ALCEP
40	74	11.3	314	1	YR83_MYCTU
41	74	11.3	427	1	THD1_MYCTU
42	74	11.3	434	1	CTPA_BARBA
43	73.5	11.3	452	1	HRA3_HUMAN
44	73.5	11.3	870	1	DLG2_HUMAN
45	73	11.2	711	1	SP3_HUMAN

ALIGNMENTS

RESULT	ID	HTRA_LACTA	STANDARD	PRT	408 AA.
AC	091A06	HTRA_LACTA			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DE	01-MAR-2002	(Rel. 41, Last annotation update)			
GN	HTRA OR LT2136				
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).				
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OX	NCBI_TaxID=1360;				
RN	(1)				
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.				
RC	STRAIN-IL1403;				
RX	MEDLINE=20177820; PubMed=10712686;				
RA	Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;				
RT	"Htra is the unique surface housekeeping protease in Lactococcus				
RT	lactis and is required for natural protein processing.";				
RL	Mol. Microbiol. 35:1042-1051(2000).				
RN	(2)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-IL1403;				
RX	MEDLINE=2125186; PubMed=11337471;				
RA	Bolotin A., Wincker P., Manger S., Jaillon O., Malarre K.;				
RT	Weissenbach J., Ehrlich S.D., Sorokin A.;				
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus				
RL	lactis sp. lactis IL1403.";				
CC	Genome Res. 11:731-753(2001).				
CC	- FUNCTION: DEGRADATION OF NATURAL PRO-PROTEIN AND FOR MATURATION OF				
CC	A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED				
CC	PROTEINS.				
CC	- SUBCELLULAR LOCATION: Membrane-bound (Probable).				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE				
CC	DEGP/DECO/DEGS FAMILY.				
CC	- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL: AF155705; AAF61294.1; -				
DR	EMBL: AEO06442; AAK06234.1; -				
DR	InterPro: IPR001478; PDZ.				
DR	InterPro: IPR001940; Protease2C.				
DR	InterPro: IPR001254; Trypsin.				
DR	Pfam: PF00595; PDZ; 1.				
DR	Pfam: PF00899; trypsin; 1.				
DR	PRINTS: PR00834; PROTEASES2C.				
DR	SMART: SM00228; PDZ; 1.				

DR PROSITE; PSS0106; PDZ; 1.
 KW Hydrolase; Serine protease; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26
 FT DOMAIN 88 284 CATALYTIC.
 FT ACT_SITE 302 383 PDZ
 FT ACT_SITE 127 127 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 157 157 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 239 239 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 408 AA; 41648 MW; 581B90B55A7DF851 CRC64;

Query Match
 Best Local Similarity 30.6%; Score 132.5; DB 1; Length 408;
 Matches 41; Conservative 18; Mismatches 44; Indels 31; Gaps 5;

QY 13 GGGFAIPGQMAIA-----GQIKLPVHIGPTAFILGLVD-----NNGN----- 53
 DB 270 GLGFAIPSDVNVINIKLADGKISRPA-----LGRVMDLSQSLTNDSSQLPSS 321
 QY 54 ---GARVORVGSAPASAGISTGDIYTVADGAPINSATAMDALNGHHPGVISYVWQT 110
 DB 322 YVGGVVVYVSGSLPAASAGLAKAGDVTYKVDPAVTSSTDLSALYSHNINDVAVTY- 380
 QY 111 KSGGRTGNVTIAE 124
 DB 381 RDKSNTADVKLK 394

RESULT 2
 HTRA_LACHE STANDARD; PRT; 413 AA.

AC 0924H7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine protease do-like htra (EC 3.4.21.-).
 GN HTRA.
 OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=53/7;
 RA MEDLINE=9047559; PubMed=9829922;
 RA Smeds A., Varmann P.K., Palva A.M.;
 RT "Molecular characterization of a stress-inducible gene from
 RT Lactobacillus helveticus."
 RL J. Bacteriol. 180:6143-6153(1998).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ005672; CAA06668.1; -
 CC MEROPS; S01.273; -
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Trypsin.
 CC Pfam; PF00595; PDZ; 1.
 CC PRINTS; PR00834; trypsin; 1.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PSS0106; PDZ; 1.
 KW Hydrolase; Serine protease; Transmembrane.
 FT TRANSMEM 20 40
 FT POTENTIAL.

FT DOMAIN 141 302 CATALYTIC.
 FT DOMAIN 305 401 PDZ.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 413 AA; 42647 MW; B16B77991C8707 CRC64;

Query Match
 Best Local Similarity 31.4%; Score 129; DB 1; Length 413;
 Matches 37; Conservative 15; Mismatches 48; Indels 18; Gaps 3;

QY 8 QLSG-----GGFAIPGQMAIA-----GQIKLPVHIGPTAFILGLVD 49
 DB 279 QSSDGTVEGMAFAIPSENVTVINELVKKGIYRPOIGRVIALDGIPIGYSRLIKS 338
 QY 50 NNGGARVORVGSAPASAGISTGDIYTVADGAPINSATAMDALNGHHPGVISYV 107
 DB 339 NLRNGIYAFVSRNGSAPASAGISTGDIYTVADGAPINSATAMDALNGHHPGVISYV 396

RESULT 3

DEGP_RHIME STANDARD; PRT; 504 AA.

ID DEGP_RHIME
 AC 052894;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DEGP1 OR DEGP OR R01021 OR SMC02365.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=96146524; PubMed=8550509;
 RA Glazebrook J., Ichige A., Walker G.C.;
 RT "Genetic analysis of Rhizobium meliloti bcaA-phoA fusion results in
 RT identification of degp: two loci required for symbiosis are closely
 RT linked to degp."
 RL J. Bacteriol. 178:745-752(1996).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=1021;
 RA MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetel D., Puchler A., Purnelle B., Ramsperger D.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U31512; AAC43669.1; ALT_INIT.
 CC MEROPS; A1591785; CAC45593.1; -
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001940; Protease2C.

```

DR InterPro:IPR000126; Ser_Protease_V8 .
DR pfam: PF000595; PDZ: 2.
DR Pfam: PR00089; trypsin: 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PM00839; V8PROTEASE.
DR SMART: SMO0228; PDz: 2.
DR ProSITE: PS50106; PDz: 2.
KM Hydrolase; Serine protease; Signal: Periplasmic; Complete proteome.
FT SIGNAL          1         26
FT CHAIN           27        504
FT DOMAIN          113       286
FT DOMAIN          287      378
FT ACT_SITE        401     491
FT ACT_SITE        140    140
FT ACT_SITE        170    170
FT ACT_SITE        244    244
FT CONFLICT         14       15
FT CONFLICT         39      147
FT FT              EDLPEDEHLRFREFAPRENDRADRRGRGEGRLP
FT FT              RAQSGFFITLEDGLVLTNNHVEDGS -> AVSPMSTPF
FT FT              RRRSSPSACRHVNASATMKATSPLISAAGSRFCRTIRCG
FT FT              YSSANSILRKVMTPYIGATAAIVAKVASVGKRAPASSSP
SQ SEQUENCE        504 AA;   53035 MW;   D7E82BB9B981EA23C CRC64;
                                KYNDSPTTSSPPRP (IN REF. 1)
                                KSADVDLYTNNAKKDKRSALFOIEAQESRYVALPIITOG
                                -> NOTSET (IN REF. 1).

Query Match                   17.5%; Score 114; DB 1; Length 504;
Best Local Similarity 30.4%; Pred. No. 0.015;
Matches 41; Conservative 16; Mismatches 58; Indels 20; Gaps 6;

OY 7 FQLSGGGG--FAIPICGMALAIGO-IKLPTVHIQPTAFI GLGVN-----NGG 52
    | | | | | | | | | | : | | | | : ||| :
Db 263 FSPSGGNVGI AFDLPASYAKDVDSL IKGDFVSRG---WLGVOIQOPTKDI AESLSSEA 319

OY 53 NGARQVRVGSAPAASLAGISITGDVTITANDGA PINSATMDALNHGHGDIVST-WQTk 111
    | | | | | : : : | | | | | : : : : | | | | | :
Db 320 NGAIVEPDQAQSPEKAIKINGVDVTA LNEGPVKDPDLRARVAALRPGSTAETVLW--R 377

OY 112 SGGRTRGTWTLAEGP 126
    | | : | | : | : |
DB 378 SKGETVNLDELICTLP 392


RESULT 4
DEGP_BRUAB STANDARD; PRT; 513 AA.
ID DEGP_BRUAB ID DEGP_BRUAB
AC Q44597;
BT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DN 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
OS DEGP OR HTRA.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bruceiaceae; Brucella.
OX NCBI_Taxid=235; [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=2308;
RX MEDLINE=95165990; PubMed=7861951;
RA Tatum F.M., Cheville N.F., Morfitt D.;
RT "Cloning, characterization and construction of htra and htra-like mutants of Brucella abortus and their survival in Balb/c mice.";
RL Microb. Pathog. 17:23-36(1994).
CC CC SUBCELLULAR LOCATION: periplasmic (Potential).
CC CC SIMILARITY BELONGS TO PEPTIDASE FAMILY SZC; ALSO KNOWN AS THE DEGP/DEGO/DEGS FAMILY.
CC CC SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: U07352; AAA70164.1; -.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR001940; Protease2C.
DR      InterPro: IPR001254; Trypsin.
DR      Pfam: PF00595; PDZ; 2.
DR      Pfam: PF00089; Trypsin; 1.
DR      PRINTS: PR00834; PROTEASES2C.
DR      SMART: SM00228; PDZ; 2.
DR      PROSITE: PS50106; PDZ; 2.
KW      Hydroxylase; Serine protease; Signal; Periplasmic.
FT      SIGNAL          1      25
FT      CHAIN           26      513
FT      DOMAIN          125      299
FT      DOMAIN          300      391
FT      DOMAIN          414      500
FT      ACT_SITE        152      152
FT      ACT_SITE        182      182
FT      ACT_SITE        257      257
FT      ACT_SITE        513 AA; 53483 MW; DELCEP1959472806 CRC64;
SQ
Query Match          16.4%; Score 107; DB 1; Length 513;
Best Local Similarity 32.6%; Pred. No. 0.059;
Matches 44; Conservative 11; Mismatches 60; Indels 20; Gaps 5.
OY      7 FOLSGGGG--FAPIGQAMAIAG-QIKLPYVHIG-----PFAPIGLGVNDNN 51
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      276 FSPSGSGVGIAIPSPSTAKQVVDLIRKGSVERGIVGOIPTYRKDIASIGLA----E 331
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      52 GNGAVQVRVGSAPASLIGISTGDTVAVDAGPINSATAMADALNHHPGVDYISVYQTR 111
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      332 EKGALVYASQDDGCPRAKAKGIRKAGDVITVANGETVDDPRDLARKVANIAGKRAALTVWRK 391
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      112 SSGTRGTNVLTAEGR 126
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      392 NKAEET-NVTIAMP 405
OY      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
RESULT      5
DEGP_SALTY
AC      DEGP_SALTY          STANDARD:          PRT;          475 AA.
AC      P26982;
DB      01-AUG-1992 (Rel. 23, Created)
DB      01-AUG-1992 (Rel. 23, Last sequence update)
DB      01-MAR-2002 (Rel. 41, Last annotation update)
DE      DEGP OR HTRA OR PTD OR STM0209.
OS      Salmonella typhimurium.
OS      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Salmonella.
OX      NCBI_TaxId=602;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C5;
RX      MEDLINE=91251770; PubMed=1645840;
RA      Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
RA      Ali T., Miller I., Hormaeche C.;
RT      "The role of a stress-response protein in Salmonella typhimurium
RT      virulence.";
RL      Mol. Microbiol. 5:401-407(1991).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX      MEDLINE=21534948; PubMed=11677609;
RA      McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA      Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
 CC SPECIFICITY WITH HHOA/DEGO.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54548; CA38420.1; -;
 DR EMBL: AE008704; AAL19173.1; -;
 DR PIR: S15337; S15337.
 DR PIR: S21327; S21327.
 DR MEROPS: S01.273; -;
 DR StyGene: SG10173; degp.
 DR Interpro: IPR001478; PDZ.
 DR Interpro: IPR001940; Protease2C.
 DR Interpro: IPR001254; trypsin.
 DR Pfam: PF00595; PDZ 1.
 DR Pfam: PF00089; trypsin. 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ 2.
 DR PROSITE: PS50106; PDZ 2.
 DR HydroLase: Serine protease: Heat shock; Periplasmic; Signal;
 KW Complete proteome.
 FT CHAIN 1 26 BY SIMILARITY.
 FT DOMAIN 27 475 PROTEASE DO.
 FT DOMAIN 281 372 PDZ 1.
 FT DOMAIN 378 467 PDZ 2.
 FT ACT_SITE 132 132 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 475 AA; 49315 MW; 86E685BF3C1A289F CRC64;
 Query Match 16.28; Score 106; DR 1; Length 475;
 Best Local Similarity 29.68; Pred. No. 0.066;
 Matches 37; Conservative 18; Mismatches 52; Indels 18; Gaps 3;
 QY 13 GGGFAIPDIGAMAIAGIKLPTVHIGPTAFGLGVNNGN-----GARVOR 59
 Db 264 GIGFAIPSNMKNLTSGM-----VEYGVKRGELGIMGTELSLAKAKVDAQRAPFSQ 319
 QY 60 VYGSAPASLIGISGDVITAVDAPINSATMAADALNGHHPGDVITSVTWQTKSGGTFRGN 119
 Db 320 VMNPNSSAKAGIRAGDVITSLNGKPISSFALRAOVGTMPGVSKISL-GLLRGKAIVN 378
 QY 120 VTIAE 124
 Db 379 LELQ 383
 RESULT 6
 DEGS_ECOLI STANDARD; PRT; 355 AA.
 AC P31137;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protease degs precursor (EC 3.4.21.-).
 GN DEGS OR HHOA OR HTRH OR B3235 OR Z4594 OR ECS4108.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; Pubmed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degp and degs, *Escherichia coli* genes encoding
 RT J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; Pubmed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; Pubmed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., DiMantia E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; Pubmed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Hal C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
 RX MEDLINE=88105815; Pubmed=332223;
 RA Vogel R.F., Entian K.-D., Mecke D.;
 RT "Cloning and sequence of the mdh structural gene of *Escherichia coli*
 RT coding for malate dehydrogenase.";
 RL Arch. Microbiol. 149:36-42(1987).
 RN [7]
 RP IDENTIFICATION.
 RA Bazan J.F., Fletcher R.J.;
 RT "Structural and catalytic models of trypsin-like viral proteases.";
 RL Semin. Virol. 1:311-322(1990).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

ID DEGP_ECOLI STANDARD: PRT; 474 AA.
 AC P09376; P15724;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR PTD OR B0161 OR 20173 OR ECS0165.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=89057448; PubMed=3057437;
 RA Lipinska B., Sharma S., Georgopoulos C.;
 RT "Sequence analysis and regulation of the htra gene of Escherichia
 coli: a sigma 32-independent mechanism of heat-inducible
 transcription.";
 RT Nucleic Acids Res. 16:10053-10067(1988).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RX MEDLINE=94261430; PubMed=8202364;
 RA Fujita N., Mori H., Yura T., Ishihama A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 2.4-4.1 min (110,917-193,643 bp) region.";
 RT Nucleic Acids Res. 22:1637-1639(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL [4]
 RN SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
 RA Lashari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RT Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Ten G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RT Nature 409:529-533(2001).
 RL [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RT DNA Res. 8:11-22(2001).
 RL [7]
 RN SEQUENCE OF 1-50 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=90323597; PubMed=2165018;

RA Onirk S., Bhattacharjee S.K., Bessman M.J.;
 RT "Primary structure of the deoxyguanosine triphosphate
 triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
 RL Gene 89:13-18(1990).
 RN [8]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE=90207273; PubMed=2157212;
 RA Wurgler S.M., Richardson C.C.;
 RT "Structure and regulation of the gene for dgt triphosphohydrolase
 from Escherichia coli.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
 RL [9]
 RN CHARACTERIZATION, AND SEQUENCE OF 27-39.
 RX MEDLINE=9020693; PubMed=2180903;
 RA Lipinska B., Zyllics M., Georgopoulos C.;
 RT "The htra (degP) protein, essential for Escherichia coli survival at
 RT high temperatures, is an endopeptidase.";
 RL J. Bacteriol. 172:1791-1797(1990).
 RN [10]
 RP IDENTITY OF HTRA AND PROTEASE DO.
 RX MEDLINE=91222240; PubMed=2025286;
 RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
 RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
 RT "Protease do is essential for survival of Escherichia coli at high
 RT temperatures: its identity with the htra gene product.";
 RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADATE
 CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
 CC -1- SUBUNIT: MULTIMERIC.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
 CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M36536; AAA23994.1; -;
 DR EMBL; X12457; CAB30997.1; -;
 DR EMBL; D26562; CAB20280.1; -;
 DR EMBL; AE000125; AAC73372.1; -;
 DR EMBL; U70214; AAB08591.1; -;
 DR EMBL; AE005192; AAG54465.1; -;
 DR EMBL; AP000250; BAB33588.1; -;
 DR EMBL; M29955; AAA23717.1; -;
 DR EMBL; M31772; AAA23680.1; -;
 DR PIR; S01899; S01899.
 DR PIR; B35993; B35993.
 DR MEROPS: S01.273; -;
 DR SWISS-2DPAGE; P09376; COLI.
 DR EcoGene; EGI0463; degP.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolase; Serine protease; Heat shock; periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 474 PROTEASE DO.
 FT DOMAIN 280 371 PDZ 1.

```

FT DOMAIN 377 466 PDZ 2.
FT ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).
FT CONFLICT 46 46 E -> Q (IN REF. 7).
FT CONFLICT 192 192 A -> G (IN REF. 1).
FT CONFLICT 467 474 STYLIMO -> RHLPLVNAVLSLPLFKTGSGPYNL (IN
FT REF. 1).
SO SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;

Query Match 15.8%; Score 103.5; DB 1; Length 474;
Best Local Similarity 32.6%; Pred. No. 0.11;
Matches 30; Conservative 13; Mismatches 32; Indels 17; Gaps 2;

OY 13 GGGFAIPGOMAINGOIKLPVHIGPFAFLGLGVNDNGN-----CARVQR 59
DB 263 GIGFAIPSNMKNLTSQM---VEYGVRKRGELGIMTELSLAKAMKVDKRGAFVSQ 318
OY 60 VVGSAPASLIGSTGDTAVDAPINSATAM 91
DB 319 VLPNSAKAGIKAGDVITSLNGKRISSPAAL 350

RESULT 9
DEGP_BARHE STANDARD; PRT; 503 AA.
AC PS4925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen htra).
GN DEGP OR HTRA.
OS Bactonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bacteriella; Bacteriella.
OX NCBI_TaxID=38523;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-HOUSTON-1;
RX MEDLINE=94299828; PubMed=8027347;
RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
RA Goral S., Hager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR.";
RL J. Clin. Microbiol. 32:942-948(1994).
CC -1- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20127; AAA97430.1; -
CC MEROPS: S01.273; -
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001940; Protease2C.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00595; PDZ; 2.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00834; PROTEASES2C.
CC SMART: SM00228; PDZ; 2.
CC PROSITE: PS50106; PDZ; 2.
CC Hydrolyase; Serine protease; periplasmic; signal.
FT SIGNAL 1 18 POTENTIAL.

```

```

FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
FT DOMAIN 286 357 LIKE.
FT DOMAIN 419 466 PDZ 1.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 503 AA; 54114 MW; 6CD9F474328AFA9E CRC64;

Query Match 15.8%; Score 103.5; DB 1; Length 503;
Best Local Similarity 27.3%; Pred. No. 0.11;
Matches 39; Conservative 15; Mismatches 52; Indels 37; Gaps 6;

OY 7 FOLISGGGCG--PAIPGOMAINGOIKLPVHIGPFAFLGLGVNDNGNGARVQV---- 60
DB 266 FSPSGGNVGIAFAIPATPAATNEVVQL-----IEKGLVCGWIGVOIQPTKEI 313
OY 61 ---VG-----SAPASLIGSTGDTAVDAPINSATFAMADALNHHGVDVI 104
DB 314 SDSIGKEAKGALITDPLKGPAAKAGIKAGDVITSVNGEKINDVRLAKRIANMSGETV 373
OY 105 SV-TWQTKSGGTRTGNVTLAEGP 126
DB 374 TLGVW--KSGKEENIKVKDSMP 394

RESULT 10
DEGP_BUCAI STANDARD; PRT; 478 AA.
AC PS7322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
DE DEGP OR BU228.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RA Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001118; BAB12943.1; -
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001940; Protease2C.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00595; PDZ; 2.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00834; PROTEASES2C.
CC SMART: SM00228; PDZ; 2.
CC PROSITE: PS50106; PDZ; 1.
CC Hydrolyase; Serine protease; signal; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 116 254 CATALYTIC.

```

FT DOMAIN 281 372 PDZ 1.
 FT DOMAIN 387 469 PDZ 2.
 FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 478 AA: 52230 MW: 868B8732CAC50629 CRC64;

Query Match 14.4%; Score 94; DB 1; Length 478;
 Best Local Similarity 25.4%; Pred. No. 0.66; Mismatches 49; Indels 18; Gaps 4;
 Matches 31; Conservative 24;

QY 13 GGGFAIPIGQAMAIAGQ-ITLPTVHIGPTAFGLGVYD-----NNGNGARYQRYVGS 63
 Db 265 GIGFAIPCMVKNKLTFRQWVQFGVGRGELGIMGELNSDLAQIMKINSOKAFYSRVLPN 324
 QY 64 APAASLGISTGVITAVDAPINSATAMADALNGHHPGDVISTWTKSGGTRTG---NV 120
 Db 325 SSAFEAGIKAGDIIITLNRKPISSEFSLRAEIGS-----LPVATKMEIGVFEGRIRKNI 378
 QY 121 TL 122
 Db 379 TV 380

RESULT 11
 DEGP_ARATH.
 ID DEGP_ARATH STANDARD; PRT: 448 AA.
 AC 09LULO:

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease Do-like 8, chloroplast precursor (EC 3.4.21.-).
 GN DEGP8 OR AT5G39830 OR K13H13.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;

RA SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RL "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
 RN Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE OF 91-113 AND 345-361.

RC STRAIN=CV, COLUMBIA;
 RA Kieselbach T., Bystedt M., Schroeder W.P.;
 RL Submitted (OCT-2000) to the SWISS-PROT data bank.
 CC -1- FUNCTION: PROBABLE SERINE PROTEASE.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID LUMEN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AB024023; BAA98101.1; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.

DR PROSITE: PS0106; PDZ; FALSE NEG.
 KW Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT TRANSIT 90 THYLAKOID.
 FT CHAIN 91 448 PROTEASE DO-LIKE 8.
 FT DOMAIN 152 333 SERINE PROTEASE.
 FT DOMAIN 336 433 PDZ.
 FT ACT_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 292 292 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 448 AA: 47492 MW: A986FC13870AFAF CRC64;

Query Match 14.1%; Score 92; DB 1; Length 448;
 Best Local Similarity 27.4%; Pred. No. 0.91; Mismatches 49; Indels 26; Gaps 5;
 Matches 34; Conservative 15;

QY 10 SQGGQGFAPITGQAMAIAGQITKLPVHIGPTAFGLGV-----VDNN---GNGARYQRYV 60
 Db 316 TSAGVGFAIPSSYVAKIVPQL-----IQFSKYLRAGINIELADPYANQINVRNGALVLOY 371
 QY 61 VGSAPASLG-----ISTGDTVAVDAGAPINSATAMADALNGHHPGDVISTWQ 109
 Db 372 PGKSLAEKAGLHPTSRGFGAGNIVLGDIIIVAVDDKPYKNKAELMKITIDEXSVGD--KVTLK 429
 QY 110 TKSG 113
 Db 430 TKRG 433

RESULT 12
 DEGP_CHLTR.
 ID DEGP_CHLTR STANDARD; PRT: 497 AA.
 AC P18584; 084630;

DT 01-NOV-1990 (Rel. 16, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
 DE immunogenic protein) (SK59).
 GN DEGP OR HTRA OR CT823.

OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;

RA SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 12;
 RX MEDLINE=90337348; PubMed=2379836;
 RA Kahan S., Weinstein Y., Sarov I.;
 RL "Cloning, characterization and sequence of a novel 59-kDa protein of
 RL Chlamydia trachomatis.";
 RN Gene 90:61-67(1990).

RP SEQUENCE FROM N.A.
 RC STRAIN=D/OW-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Iammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RL "Genome sequence of an obligate intracellular pathogen of humans:
 RL Chlamydia trachomatis.";
 RL Science 282:754-759(1998).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
 CC ON THE AUTHORS TRANSLATED THEIR PUTATIVE 59 KDA IMMUNOGENIC PROTEIN
 CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
 CC PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way


```

RT      Influenza Rd.":
RL      Science 269:496-512(1995).
CC      -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC      -1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
CC      (PROTEASE DO) AND HHOA.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC      DEGR/DEGRQ/DEGS FAMILY.
CC      -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.slb.ch/announce/
CC      or send an email to license@isb.slb.ch).
CC      -----
DR      EMBL; U32805; AAC22906.1; -.
DR      MEROPS; S01.274; -.
DR      TIGR; H11259; -.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR001940; Protease2C.
DR      InterPro: IPR001254; Trypsin.
DR      Pfam; PF00595; PDZ; 2.
DR      Pfam; PF00089; trypsin; 1.
DR      PRINTS; PR00834; PROTEASES2C.
DR      SMART; SM00228; PDZ; 2.
DR      PROSITE; PS0106; PDZ; 2.
KW      Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT      SIGNAL          1      29
FT      CHAIN           30      466
FT                                     PROBABLE PERIPLASMIC SERINE PROTEASE
FT                                     DO/HHOA-LIKE.
FT      DOMAIN          270      361
FT      FT              367      458
FT      ACT SITE        120      120
FT      ACT SITE        150      150
FT      ACT SITE        226      226
FT      SEQUENCE        466 AA; 49434 MW; ED050A00047B5851 CRC64;
SQ
Query Match          13.4%; Score 87.5; DB 1; Length 466;
Best Local Similarity 28.4%; Pred. No. 2.2;
Matches 31; Conservative 16; Mismatches 45; Indels 17; Gaps 3;
OY      13 GGGFAPLPGQAMATAGQIKLPVHIGPP-AFLGLGVVDNNGN-----GARVQR 59
DB      253 GIAEFAIPSNQSNMLVQQT-----LEFGVRRKGLGTIGKGLNADLAKAFNVAQSGAFVSE 308
OY      60 VVGSAPASLIGSTGDTTAVADGAPINATFATMAADALNCHHPGDVYSVTV 108
DB      309 VLPKSAARAKAGLAKGDDITTANNGOKTISFAETRAKIAITTGAKKEISLTV 357
RESULT 15
MURF_MYCTU ID MURF_MYCTU STANDARD; PRT; 510 AA.
AC 006220;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylglucosaminyl-1,4-D-glucanase (UDP-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNac-pentapeptide synthetase)
GN MURF OR RV2157C OR MT2216 OR MTC270.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RX MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Dawscock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RP [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fritschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson S.J., Deboy R., Dodson R., Gwinn M.T., Haft D., Hickey E.,
RA Kolner J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A.L., Utterback T., Feldman J., Khouri H., Gill J., Mikulic A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetyluramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanoate + D-alanyl-D-alanine - ADP
CC + phosphate + UDP-N-acetyluramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE MURCODE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; Z95388; CAB08670.1; -;
DR EMBL; AE007068; AAK6500.1; -;
DR HSSP; P11880; 16G4.
DR TIGR; MT2216; -;
DR Tuberculist; RV2157c; -;
DR InterPro; IPR000713; Mur_Ligase.
DR InterPro; IPR004101; Mur_Ligase_C.
DR Pfam; PF01225; Mur_Ligase; 1.
DR Pfam; PF02875; Mur_Ligase_C; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW ATP-binding; Complete proteome.
NP_BIND 136 142 ATP (POTENTIAL).
FT CONFLICT 49 49 A -> R (IN REF. 2).
FT CONFLICT 442 442 G -> E (IN REF. 2).
FT SEQUENCE 510 AA; 51632 MW; 7BE83EDD1BAA2BEC4 CRC64;

Query Match 13.2%; Score 86; DB 1; Length 510;
Best Local Similarity 26.3%; Pred. No. 3.3;
Matches 46; Conservative 24; Mismatches 57; Indels 48; Gaps 11;

QY 1 TAAADNFDLSGGGCGFAIP-----IGQMAIAGQ- IKIPVHIGPTA-- 41
DB 33 TVEEDSRAIGGGFLAIPGARADGHDHMAASAVAAGAAVVAAPVAPVAPVAPN 92
QY 42 -FLGAGVVDNNGNCAVQRYV---VGSAPPAAL-----GI-----STGDDITAVDG 82
DB 93 VLAGVLEHDNDGSGAAVLAALAKLATATAVAADLVAGGLTIIGTISGCKSTKIDMAAV-L 151
QY 83 APIASATAMADALNGH--HPGDVTSVTWQ-----KSGGRTGNV-TLAE-GPPA 128
DB 152 APIDEVVAAPSSPNNELGHPVTVLRAFRRTDYLITEMAARHNGITALAEAPRS 206

Tue Aug 13 10:27:59 2002

us-09-684-215a-18.rsp

Page 11

Search completed: August 13, 2002, 10:42:46
Job time: 389 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 10:42:23 ; Search time 43.59 Seconds
(Without alignments)
507.992 Million cell updates/sec

Title: US-09-684-215a-18
Perfect score: 653
Sequence: 1 TRASNFOLOGGCGFAIP1.....QTRSGGTRGNVTLAEGPPA 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mmc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632	96.8	355	16	007175 mycobacteri
2	437.5	67.0	361	2	050320 mycobacteri
3	422.5	64.7	354	16	09C9Y mycobacteri
4	200.5	30.7	464	16	053896 mycobacteri
5	185.5	28.4	382	16	09CDB7 mycobacteri
6	185.5	28.4	452	2	092566 mycobacteri
7	163.5	25.0	542	2	09FBK9 streptomyce
8	149	22.8	519	2	09J3J0 streptomyce
9	134.5	20.6	362	2	09S2R5 streptomyce
10	128.5	19.7	474	2	09ALST pseudomonas
11	123.5	18.9	371	2	031388 bradyrhizob
12	123.5	18.9	474	16	057155 pseudomonas
13	122.5	18.8	514	16	09PBA3 xyella fas
14	122	18.7	394	16	P72780 synechocyst
15	120	18.4	473	2	044476 azotobacter
16	119	18.2	441	16	Q9RTK4 delinococcus

17	116.5	17.8	203	2	Q9X840 streptomyce
18	115	17.6	348	16	Q97L01 clostridium
19	114	17.5	513	16	Q98CS8 rhizobium l
20	112.5	17.2	530	16	Q9A4S2 caulobacter
21	111	17.0	499	16	Q9JVT1 neisseria m
22	110	16.8	481	2	Q9A0D1 pseudomonas
23	110	16.8	516	16	Q985F9 rhizobium l
24	109.5	16.8	465	16	Q920E6 rhizobium m
25	109	16.7	393	16	Q97N37 streptococc
26	109	16.7	397	2	Q06670 streptococc
27	106.5	16.3	459	16	Q9WZ41 thermotoga
28	106	16.2	500	2	Q44652 bruceella ab
29	105.5	16.2	452	16	P73354 synechocyst
30	105.5	16.2	491	2	Q9LBR0 shigella so
31	103.5	15.8	407	16	Q99XG9 streptococc
32	103	15.8	397	2	Q33F87 streptococc
33	102.5	15.7	404	16	Q3351 mycobacteri
34	102	15.6	472	16	Q9A8R9 caulobacter
35	102	15.6	476	16	Q9ZM18 helicobacte
36	101	15.5	398	16	Q83752 treponema p
37	100.5	15.4	453	2	Q9FD11 aeromonas h
38	100	15.3	500	2	Q9KJN6 myxococcus
39	98.5	15.1	481	16	Q9PGL3 xyella fas
40	97.5	14.9	428	16	Q98N31 rhizobium l
41	97.5	14.9	474	2	Q44596 bacillus su
42	96.5	14.8	224	16	Q35021 bacillus su
43	96.5	14.8	458	2	Q9R911 bacillus su
44	95.5	14.6	433	16	Q97GD5 clostridium
45	95.5	14.6	453	16	Q67436 aquifex aeo

ALIGNMENTS

RESULT 1
007175 PRELIMINARY; PRT; 355 AA.
AC 007175;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 34.9 KDA PROTEIN.
GN PEPA OR RV0125 OR MRC1A18B.07.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230; Garnier T., Churcher C., Harris D., Cole S.T., Brosch R., Parkhill J., Garnier T., Barry C.E. III, Tekala F., Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Hamlin N., Holtroyd S., Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Skelton S., Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544(1998).
RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL; Z66071; CAB09453.1; -
DR MEROPS; S01.00C; -
DR Tuberculin; RV0125; -
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; Trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN.1.
 KW Complete proteome; Hydrolase; Hypothetical protein; Serine protease.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.8%; Score 632; DB 16; Length 355;
 Best Local Similarity 95.5%; Pred. No. 3.6e-41;
 Matches 126; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVDDNNGNGAR 56
 DB 224 TAASDNFOLSGGGGFALPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVDDNNGNGAR 283
 QY 57 VORVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVYSVTWQTKSGGTR 116
 DB 284 VORVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVYSVTWQTKSGGTR 343
 QY 117 TGNVTLAEGPPA 128
 DB 344 TGNVTLAEGPPA 355

RESULT 2
 ID 050320 PRELIMINARY; PRT; 361 AA.
 AC 050320;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE 34KDA PROTEIN PRECURSOR.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JDB8/107.
 RX MEDLINE=95005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982(1994).
 DR EMBL: Z230932; CAA80638.1; -
 DR MEROPS; S01.0PC; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease; Signal.
 FT SIGNAL 1
 FT SIGNAL 38
 FT SIGNAL 38
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 67.0%; Score 437.5; DB 2; Length 361;
 Best Local Similarity 65.9%; Pred. No. 3.4e-26;
 Matches 87; Conservative 18; Mismatches 22; Indels 5; Gaps 2;

QY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVDDNNGNGAR 56
 DB 231 TAATDSYKMS-GGGGFALPIGRMAVANQIRSGAGSNTVHIGPTAFGLGVDDNNGNGAR 289
 QY 57 VORVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVYSVTWQTKSGGTR 116
 DB 290 VORVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVYSVTWQTKSGGTR 349
 QY 117 TGNVTLAEGPPA 128
 DB 350 TANTTLAEGPPA 361

RESULT 3
 ID 09CCY9 PRELIMINARY; PRT; 354 AA.
 AC 09CCY9;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PROBABLE SECRETED SERINE PROTEASE.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1769;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Bigmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felkell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC 1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL583926; CAC32191.1; -
 DR MEROPS; S01.0PC; -
 DR Leproma; ML2659; -
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR00126; Ser-Proteas_V8.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR PRINTS; PR00839; VBPROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EAAA CRC64;

Query Match 64.7%; Score 422.5; DB 16; Length 354;
 Best Local Similarity 62.9%; Pred. No. 4.7e-25;
 Matches 83; Conservative 18; Mismatches 26; Indels 5; Gaps 2;

QY 1 TAASDNFOLSGGGGFALPIGQAMAIAGQIK----LPTVHIGPTAFGLGVDDNNGNGAR 56
 DB 224 TAATDSYKMS-LGGGFALPIGQAEVVGAIIRSGAGSNTVHIGPTAFGLGVDDNNGNGAR 282
 QY 57 VORVVGSAAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVYSVTWQTKSGGTR 116
 DB 283 VAVVATGPAAMAGISGDIITVSDGVPISEATAMTVLVIPHHGELVAVVYRSAGGDL 342
 QY 117 TGNVTLAEGPPA 128
 DB 343 TANTTLAEGPPA 354

RESULT 4
 ID 053896 PRELIMINARY; PRT; 464 AA.
 AC 053896;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE SERINE PROTEASE.
 GN RV0983 OR MT0044.11.

RA RA_Murphy_L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutherford S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -;
DR MEROPS; S01.0PC; -;
DR Leptoma; ML0176; -;
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF000595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASE2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; hydrolase; Protease; Serine protease.
KW SEQUENCE 382 AA; 37084 MW; 3DD8DDBAE32A80D CRC64;

Query Match	28.48;	Score 185.5;	DB 16;	Length 382;
Best Local Similarity	38.18;	Pred. No. 8.9e-07;		
Matches 48;	Conservative 18;	Mismatches 53;	Indels 7;	Gaps 2;

```
QY      119 NVTLAE 124
      III :
Db      374 QVTLGK 379
```

```

RESULT      6
ID          0925G6      PRELIMINARY;      PRT;      452 AA.
AC          0925G6;
DT          01-MAY-1999 (TREMBLrel. 10; Created)
DT          01-MAY-1999 (TREMBLrel. 10; Last sequence update)
DE          01-DEC-2001 (TREMBLrel. 19; Last annotation update)
ID          PUTATIVE SERINE PROTEASE.
GN          MCB373.28.
OS          Mycobacterium leprae.
OC          Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX          Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN          [1]
RP          SEQUENCE FROM N.A.
RA          Harris D.; Taylor K.;
RL          Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN          [2]
RP          SEQUENCE FROM N.A.
RA          James K.D., Parkhill J., Barrell B.G., Randsream M.A.;
RL          Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN          [3]
RP          SEQUENCE FROM N.A.
RX          MEDLINE=93188700; PubMed=8446027;
RA          Elgimeleer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT          "Use of an ordered cosmid library to deduce the genomic organization
RT          of Mycobacterium leprae.";
RL          Mol. Microbiol. 7:197-206(1993).
DR          EMBL; AL035500; CAB36690; -.
DR          MEROPS: S01.UPC; -.
DR          InterPro: IPR001478; PDI.
DR          InterPro: IPR001940; Protease2C.
DR          InterPro: IPR001254; Trypsin.
DR          Pfam: PF00595; PDZ; 1.

```

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB091F983 CRC64;

Query Match 28.4%; Score 185.5; DB 2; Length 452;
 Best Local Similarity 38.1%; Pred No. 1.1e-06;
 Matches 48; Conservative 18; Mismatches 53; Indels 7; Gaps 2;

QY 2 AASDFOLSGGGGFAIPGQAMAIAGQIKLPVTHIGPTAFGLGV---VDNNGGARVQ 58
 DB 328 ADSGDAGSGSIGLGFALPVDQAKRIADEL---ISTGKATFASLGVQVATMDKCTPGAKVM 383
 QY 59 RVGSGAPASLGISTGVDYITAVDGAIPNSATMADALNGHHPGDVTSVTWQKSGGTRTG 118
 DB 384 DVVAGGAANAAVAPKGVVLTGVDRILISSDALVAVRKAGDKVSLTYODSGSSRTV 443
 QY 119 NVTLAE 124
 DB 444 QVTICK 449

RESULT 7
 Q9FBK9 PRELIMINARY; PRT; 542 AA.
 AC 09FBK9;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DE PUTATIVE PROTEASE.
 GN SCP8.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN 11
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RA Saunders D., Harris D.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN 12
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN 13
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL390975; CAC01350.1; -;
 DR MEROPS; S01.UPC; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR002965; P-rich_tensn.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00059; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR01217; PRICHTEXTNSN.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Protease; serine protease.
 SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

Query Match 25.0%; Score 163.5; DB 2; Length 542;
 Best Local Similarity 34.6%; Pred. No. 6.5e-05;
 Matches 46; Conservative 17; Mismatches 15; Indels 15; Gaps 4;

QY 4 SDNFOLSGGGGFAIPGQAMAIAGQIKLPVTHIGPTAFGLGV---VDNN--GNGARVQ-- 58
 DB 412 SDDGAGSIGLGFALPINOGRVAEEL-----INTGKAHPVYGITLDANTYDGAIVSAK 467
 QY 59 -----RVGSGAPASLGISTGVDYITAVDGAIPNSATMADALNGHHPGDVTSVTWQKS 112
 DB 468 GGDGPVATVTGPGAKAGIKPGVDYITAVDQGVHSGEELIYKTRAHRGDRLETLRLDRG 527
 QY 113 GGRFTGVNTLAEG 125
 DB 528 KETVSLVLSGS 540

RESULT 8
 Q9J3J0 PRELIMINARY; PRT; 519 AA.
 AC 09J3J0;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DE PUTATIVE PROTEASE.
 GN SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN 11
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN 12
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN 13
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL596251; CAC44701.1; -;
 KW Protease.
 SQ SEQUENCE 519 AA; 50327 MW; 5CBBD5F0CC19E428 CRC64;

Query Match 22.8%; Score 149; DB 2; Length 519;
 Best Local Similarity 33.9%; Pred. No. 0.00081;
 Matches 41; Conservative 20; Mismatches 44; Indels 16; Gaps 5;

QY 13 GCGFAIPGQAMAIAGQIKLPVTHIGPTAFGLGV---VDNNGGARV--ORVVGSG-- 63
 DB 397 GIGFAIPVNOAKFVAOOL---IKSGPVYAKIGASVSLBETTGAKLTGEGVGSDPVE 452
 QY 64 --AASLGISTGVDYITAVDGAIPNSATMADALNGHHPGDVTSVTWQKSGGTRGNT 121
 DB 453 KGGPADAGLKPQDVITKLDRAVIDSGPTLIGETIWHKPGDEVTVYE--KGGKHTAEVT 511
 QY 122 L 122
 DB 512 L 512

RESULT 9

```

Q9S2R5      Q9S2R5      PRELIMINARY;      PRT;      362 AA.
AC      Q9S2R5;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      POTATIVE SECRETED PROTEASE.
GN      SC5F7.30.
OS      Streptomyces coelicolor.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=1902;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RA      Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL      Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RA      Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL      Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RA      MEDLINE=9700351; PubMed=8843436;
RA      Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA      Khashi H., Hopwood D.A.;
RT      A set of ordered cosmids and a detailed genetic and physical map for
RT      the 8 Mb Streptomyces coelicolor A3(2) chromosome.;
RL      M01. Microbiol. 21:77-96(1996).
DR      EMBL; AL096872; CAB51255.1; -.
DR      MEROPS; S01.0PC; -.
DR      InterPro; IPR001478; PDZ.
DR      InterPro; IPR001940; Protease2C.
DR      InterPro; IPR001254; Trypsin.
DR      Pfam; PF00595; PDZ.1.
DR      Pfam; PF00089; trypsin.1.
DR      PRINTS; PR00834; PROTEASES2C.
DR      SMART; SM00228; PDZ.1.
DR      PROSITE; PS50106; PDZ.1.
DR      PROSITE; PS50240; TRYPSIN_DOM.1.
KW      Hydrolase; Protease; Serine protease.
SQ      SEQUENCE 362 AA; 36381 MW; 71CC41F943A6D03B CRC64;

Query Match      20.6%; Score 134.5; DB 2; Length 362;
Best Local Similarity 35.2%; Pred. NO. 0.007;
Matches 43; Conservative 14; Mismatches 50; Indels 15; Gaps 5;

QY      13 GGGFAIPIGQAMAIAGQIKLPTVHIGPT----AFLGL--GVNNGN--GARVQRVVG 62
AC      09ALSI      PRELIMINARY;      PRT;      474 AA.
AC      09ALSI;
DT      01-JUN-2001 (TREMBLrel. 17, Created)
DT      01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      MUCD.
GN      MUCD.
OS      Pseudomonas aeruginosa.

```

```

OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC      Pseudomonas.
OX      NCBI_TaxID=287;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-UCBPP-PA14;
RA      Torrey P.S., Raimo L.G., Tan M., Ausubel F.M.;
RT      "The Roles of mucD and Alginate in the Virulence of Pseudomonas
RT      aeruginosa in Plants, Nematodes, and Mice.";
RL      Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF343973; AAK11276.1; -.
DR      InterPro; IPR001478; PDZ.
DR      InterPro; IPR001940; Protease2C.
DR      InterPro; IPR001254; Trypsin.
DR      Pfam; PF00595; PDZ.2.
DR      Pfam; PF00089; trypsin.1.
DR      PRINTS; PR00834; PROTEASES2C.
DR      SMART; SM00228; PDZ.2.
DR      PROSITE; PS50106; PDZ.1.
DR      PROSITE; PS50240; TRYPSIN_DOM.1.
KW      Hydrolase; Serine protease.
SQ      SEQUENCE 474 AA; 50349 MW; A22FD4338B859D4C CRC64;

Query Match      19.7%; Score 128.5; DB 2; Length 474;
Best Local Similarity 32.4%; Pred. NO. 0.028;
Matches 44; Conservative 16; Mismatches 53; Indels 23; Gaps 4;

QY      4 SDNFDLSG--GCGFAIPIGQAMAIAGQIKLPTVHIGPTAFLGLGVNNGN----- 53
DB      233 SGIETRRSGFMGLSFALPIDVALNADQK----KAGKVSRLMLGVIGEVKMDLAESFG 288
QY      54 -----GARVQRVVGSAPASLGISTGDDVTAVDGAIPNSATMADALNGHNGDVISV-- 106
DB      289 LDKPSGALVAQLVEDGPAKAGGIQVGDVTLISNGOSINESADLPHLVGNMKBGDKINLDV 348
QY      107 --TWQTKSGGTRTGNV 120
DB      349 INNGQRKSLMAVGNL 364

RESULT 11
031388      PRELIMINARY;      PRT;      371 AA.
AC      031388;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      DEGP PROTEIN.
GN      DEGP.
OS      Bradyrhizobium japonicum.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Bradyrhizobium group; Bradyrhizobium.
OX      NCBI_TaxID=375;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=98114461; PubMed=9446679;
RA      Narderahaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
RT      "Identification of the Bradyrhizobium japonicum degp gene as part of
RT      an operon containing small heat shock protein genes.";
RL      Arch. Microbiol. 169:89-97(1998).
DR      EMBL; Y13616; CAA73938.1; -.
DR      HSP; P31016; IBEF.
DR      InterPro; IPR001478; PDZ.
DR      InterPro; IPR001940; Protease2C.
DR      InterPro; IPR001254; Trypsin.
DR      Pfam; PF00595; PDZ.1.
DR      Pfam; PF00089; trypsin.1.
DR      PRINTS; PR00834; PROTEASES2C.
DR      SMART; SM00228; PDZ.1.
DR      PROSITE; PS50106; PDZ.1.
DR      PROSITE; PS50240; TRYPSIN_DOM.1.
KW      Hydrolase; Serine protease.

```


RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M.M., Meidanis J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RU Nature 406:151-159(2000).
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AE004037; AAF85040.1; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001264; Ser-proteas_V8.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00395; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0834; PROTEASES2C.
 DR PRINTS: PRO0839; V8PROTEASE.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Serine protease.
 SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;

Query Match 18.8%; Score 122.5; DB 16; Length 514;
 Best Local Similarity 36.9%; Pred. No. 0.088;
 Matches 41; Conservative 13; Mismatches 42; Indels 15; Gaps 4;

QY 4 SDFNLSQG--GQGFALPIGQAMAIAGQIK-----LPVHIGPTAFL--GIGVVDN 50
 DB 262 SQIFSGAGVGWISFALPIPINALMAEDIRTKGVQRBSMLGVEIGFIDALKAGGIGLDPDS 321
 QY 51 NGNCARQVRVGSAPASLSIGSTDVITAVDGPINSATAMADALNGHHHPG 101
 DB 322 --RGALVNNIPHPSPAKAGIEVDVIRSVNGKVISFSFDLPILIGMPPG 370

RESULT 14
 P72780 PRELIMINARY; PRT; 394 AA.
 AC 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE PROTEASE HHOA.
 GN HHOA OR SLI1679.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_Taxid=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirotsawa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Matanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions,"
 RL DNA Res. 3:109-136(1996).
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: D90900; BAA16795.1; -.
 DR MEROPS: S01.274; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00395; PDZ; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.

DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Complete proteome; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;

Query Match 18.7%; Score 122; DB 16; Length 394;
 Best Local Similarity 32.8%; Pred. No. 0.071;
 Matches 43; Conservative 11; Mismatches 39; Indels 38; Gaps 5;

QY 13 GQGFALPIGQAMAIAGQIKLPVHIGPTAFLGIGVVD-----NGN----- 53
 DB 261 GIGFALPIDQAKAI--QNTLAGGTVPHPYIGVQMMNTITVDQAOQNNRNPSPFIPEVD 318
 QY 54 GARQVRVGSAPASLSIGSTDVITAVDGPINSATAM-----ADALNGHH 99
 DB 319 GILVMRVLPGPAPRAGIRKRDVIVAVDGTPLSDGARLQRIVEAGLKNALKLIDLRGR 378
 QY 100 PGDIVSVTWQT 110
 DB 379 R--LSLTVO 386

RESULT 15
 Q44476
 ID Q44476 PRELIMINARY; PRT; 473 AA.
 AC Q44476;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MUCD.
 OS Azotobacter vinelandii.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Azotobacter.
 OX NCBI_Taxid=354;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-UW 136.
 RA MEDLINE-96178940; PubMed=8606151;
 RA Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,
 RA Soberon-Chavez G., Deretic V.;
 RT "Characterization of the genes coding for the putative sigma factor
 RT Algu and its regulators MucA, MucB, MucC, and MucD in Azotobacter
 RT vinelandii and evaluation of their roles in alginate biosynthesis,"
 RL J. Bacteriol. 178:1800-1808(1996).
 DR EMBL: U30799; AAB01513.1; -.
 DR MEROPS: S01.UNC; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00395; PDZ; 2.
 DR Pfam: PF00089; trypsin; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 473 AA; 50282 MW; C97B357D897738AB CRC64;

Query Match 18.4%; Score 120; DB 2; Length 473;
 Best Local Similarity 31.2%; Pred. No. 0.12;
 Matches 43; Conservative 16; Mismatches 59; Indels 20; Gaps 4;

QY 4 SDFNLSQG--GQGFALPIGQAMAIAGQIKLPVHIGPTAFLGIGVVDNNGN----- 53
 DB 233 SQIFRSGGFMGISFALPIEVANGVADOLKA---TGKVARGMVIGVIGEVNKRDLAESFG 288
 QY 54 -----GARQVRVGSAPASLSIGSTDVITAVDGPINSATAMADALNGHHHPGDIVSVTW 108
 DB 289 LDRPAGALVAQVLEDSRPAKGGIQQVGDVILSLDGNPIVMSADLPHLYGKLRGAANLE- 347
 QY 109 QTKSGGTRGAVTLAEGP 126
 DB 348 VVRDGKRRNIATVGAIP 365

Tue Aug 13 10:28:00 2002

us-09-684-215a-18.rpt

Page 8

Search completed: August 13, 2002, 10:42:25
Job time: 398 sec
